THE BULLETIN OF Mathematical BIOPHYSICS

SEPTEMBER 1950

THE UNIVERSITY OF CHICAGO PRESS · CHICAGO · ILLINOIS

VOLUME 12 · · · · · NUMBER 3

THE BULLETIN OF MATHEMATICAL BIOPHYSICS

EDITED BY N. RASHEVSKY

The Bulletin is devoted to publications of research in Mathematical Biology, as described on the inside back cover.

THE BULLETIN is published by the University of Chicago at the University of Chicago Press, 5750 Ellis Avenue, Chicago 37, Illinois, quarterly, in March, June, September, December. ¶The subscription price is $7.50 per year, the price of single copies is $2.25. Orders for service of less than a full year will be charged at the single-copy rate. ¶Patrons are requested to make all remittances payable to The University of Chicago Press in postal or express money orders or bank drafts.

THE FOLLOWING is an authorized agent:

For the British Empire, except North America, India, and Australasia: The Cambridge University Press, Bentley House, 200 Euston Road, London, N.W. 1. Prices of yearly subscriptions and of single copies may be had on application.

CLAIMS FOR MISSING NUMBERS should be made within the month following the regular month of publication. The publishers expect to supply missing numbers free only when losses have been sustained in transit, and when the reserve stock will permit.

BUSINESS CORRESPONDENCE should be addressed to The University of Chicago Press, Chicago 37, Ill.

COMMUNICATIONS FOR THE EDITOR and manuscripts should be addressed to N. Rashevsky, Editorial Office of *The Bulletin of Mathematical Biophysics*, 5741 Drexel Avenue, Chicago 37, Ill.

IMPORTANT ANNOUNCEMENT

We are pleased to announce that the *Bulletin of Mathematical Biophysics* will now be available in microfilm form. Complete *Bulletin* volumes may be obtained in a single roll of positive microfilm on adequately labeled metal reels at a cost of approximately one-fourth of a cent per page, which is about equal to that of preserving them in conventional library binding. Sales will be restricted to those subscribing to the paper edition, and the film copy will be distributed only at the completion of the volume year, which ends with the December issue. Inquiries concerning purchase should be directed to University Microfilms, 313 North First Street, Ann Arbor, Michigan.

PRINTED BY THE DENTAN PRINTING COMPANY . . . COLORADO SPRINGS, COLORADO

BULLETIN OF
MATHEMATICAL BIOPHYSICS
VOLUME 12, 1950

MATHEMATICAL BIOLOGY OF SOCIAL BEHAVIOR: IV. IMITATION EFFECTS AS A FUNCTION OF DISTANCE

N. RASHEVSKY
COMMITTEE ON MATHEMATICAL BIOLOGY
THE UNIVERSITY OF CHICAGO

In previous publications, social groups have been studied in which each individual has a preference for one of two possible mutually exclusive activities. This preference is measured by a quantity ϕ. The value $\phi = 0$ corresponds to no preference; a preference for one activity is measured by a positive ϕ, the preference for the other by a negative ϕ. The quantity ϕ varies from individual to individual. It has been shown previously that, owing to effects of imitation, even when the average ϕ for the group is zero, one of the two behaviors will be chosen by the majority of the group. Whereas in previous studies the imitation effect was considered as independent of the distance between the imitating and imitated individuals, in the present study the case is considered in which the effect of imitation decreases with the distance between the individuals. It is found that under certain conditions a greater percentage near the center of the area occupied by the group, rather than near the periphery, exhibits the chosen behavior. The possible sociological meaning of this gradient of behavior is discussed.

This paper is a generalization of previous studies (Rashevsky, 1949; 1950) and familiarity of the reader with either one of the above references is essential. In those previous publications we considered the effect of imitation of one individual by another as being independent of the physical distance between the two individuals. Here we shall consider the more general case, in which imitation is a function of distance. The most direct cause of the dependence of imitation upon distance may lie in the circumstance that the farther away the two individuals are geographically, the less frequently they are likely to see each other. In a properly developed theory, the function which determines the dependence of imitation on distance should be derived from such considerations. It will involve as parameters such quantities as the physical mobility of the individuals, which depends on methods of transportation, and the amount of communication transmissible from one individual to another, as a function of their distance. Leaving these possibilities to the future, we shall in-

vestigate here a very special case, in which a function of distance is chosen so as to make our equations soluble. This at least shows some general possible properties of such situations.

We shall use the same notations as in *loc. cit.*, except that we shall denote now by x and y the geographic coordinates of an individual. For the number of those individuals with a given ϕ who exhibit respectively behaviors R_1 and R_2 we shall now use $n(\phi)$ and $m(\phi)$ respectively. We shall assume a constant density of population.

If the imitation depends on the distance, then, in general, the quantity ψ will be a function of the coordinates x, y of the individual. Denote by $X(x', y')\ dx'\ dy'$ the number of those individuals whose coordinates lie in the interval $(x',x' + dx')$ and $(y',y' + dy')$ and who exhibit behavior R_1. Denote by $Y(x',y')\ dx'\ dy'$ the corresponding quantity for behavior R_2. Let $K(x,y;\ x',y') = K(r_{12}) = K[(x - x'^2 + (y - y')^2]$ be a function of the distance

$$r_{12} = \sqrt{(x-x')^2 + (y-y')^2}$$

between the two points x,y and x',y'. The function $K(r_{12})$ is symmetric with respect to the points x,y and x',y', so that

$$K(x,y;\ x',y') = K(x',y';\ x,y). \tag{1}$$

Generalizing the concept used in *loc. cit.* we consider that the quantity

$$[X(x',y') - Y(x',y')]\ K(x,y;\ x',y')\,dx'\ dy' \tag{2}$$

contributes to the stimulus E which results in an increase of ψ for individuals located at the point x,y. The total stimulus E is given by

$$E = \int_S \int [X(x',y') - Y(x',y')]\ K(x,y;\ x',y')\ dx'\ dy', \tag{3}$$

where the integration is extended over the whole area S occupied by the population. Therefore, instead of equation (8) of *loc. cit.*, we now obtain

$$\frac{d\psi(x,y)}{dt} = A \int_S \int [X(x',y') - Y(x',y')]\ K(x,y;\ x',y')\ dx'\ dy' - a\psi(x,y). \tag{4}$$

By the same argument as that used in *loc. cit.* we find that $X(x',y') - Y(x',y')$ is a function of $F(\psi)$ of $\psi(x',y')$ at the point x',y'. Depending on the assumptions which we make about the distribution function

for ϕ and for ξ, we find $F(\psi)$ to be either of the form given in *loc. cit.*, by H. G. Landau (1950), or H. D. Landahl (1950). In all cases, it is a function which is zero for $\psi = 0$, then for $\psi > 0$ increases monotonically to the asymptotic value $\bar{N}_0(x',y')$, where $\bar{N}_0(x',y')$ is the population density at x',y'. Hence equation (4) may be written

$$\frac{d\psi(x,y)}{dt} = A \int_S \int F[\psi(x',y')]\, K(x,y;x',y')\, dx'dy' - a\psi(x,y)\,. \quad (5)$$

The solution of this nonlinear integrodifferential equation depends on the size and shape of the area S. We shall limit ourselves here to the very special case in which the area S is a circle of radius R, and in which all quantities are functions only of the distance r from center. Furthermore, we shall restrict ourselves to the study of the steady states, in which $d\psi/dt = 0$. Equation (5) now becomes

$$a\psi(r) = A \int_0^r F[\psi(r')]\, K_1(r,r')\, dr'\,. \quad (6)$$

For the function $K[(x - x')^2 + (y - y')^2]$ we may choose any plausible function of r_{12}, which decreases with distance, tending to zero when $r_{12} \to \infty$, and remains always positive. To make our equations soluble we shall put as an approximation

$$K(x,y;x',y') = 1 - \mu[(x-x')^2 + (y-y')^2]\,, \quad (7)$$

that is, use an inverted parabola. For sufficiently large distances K becomes negative, which is physically meaningless. If, however, we take

$$\mu < \frac{1}{4R^2} \quad (8)$$

we insure the positivity of K for all pairs of points within the circle of radius R. The inverted parabola may be considered as an approximation, for not too great a distance, to a curve which, for example, looks like a normal distribution curve, and approaches zero asymptotically.

Introducing polar coordinates r and θ, we have now for any function $f(x',y')$

$$\int\int_S f(x',y')\ K(x,y;x',y')\ dx'\ dy' = \int\int_S f(x',y')\,[1$$

$$-\mu(x-x')^2+(y-y')^2]\ dx'\ dy' \tag{9}$$

$$= \int_0^R \int_0^{2\pi} f(r',\theta')\,[1-\mu(r^2+r'^2-2rr'\cos(\theta-\theta')]\ r'\ dr'\ d\theta'.$$

If f does not depend on θ, so that $f(r',\theta') = f(r')$, then expression (9) becomes, after integrating with respect to θ'

$$2\pi\int_0^R f(r)\,[1-\mu(r^2+r'^2)]\ r'\ dr'. \tag{10}$$

We thus see that in equation (6) $K_1(r,r') = 2\pi[1-\mu(r^2+r'^2)]\ r'$. Putting

$$\frac{2\pi A}{a} = \lambda, \tag{11}$$

we may write equation (6) thus:

$$\psi(r) = \lambda\int_0^R F[\psi(r')]\,[1-\mu(r^2+r'^2)]\ r'\ dr'. \tag{12}$$

Since $F(0) = 0$, therefore $\psi = 0$ is a solution of this integral equation. We shall now look for other possible solutions.

Equation (12) may also be written

$$\psi(r) = \lambda(1-\mu r^2)\int_0^R F[\psi(r')]\ r'\ dr' - \lambda\mu\int_0^R F[\psi(r')]\ r'^3\ dr'. \tag{13}$$

Both integrals are constants, independent of r. Putting

$$L = \lambda\int_0^R F(\psi(r))\ rdr; \quad M = \lambda\mu\int_0^R F[\psi(r)]\ r^3\ dr, \tag{14}$$

we find from (13)

$$\psi(r) = L - M - \mu L r^2. \tag{15}$$

Since $F(\psi)$ is a known function, therefore by introducing (15) into (14), we obtain two equations for the determination of the two constants L and M. If those equations have real roots, then (15) represents a solution of the integral equation (12).

To find the two equations, we introduce ψ as the integration variable in (14). We have from (15)

$$r\,dr = -\frac{d\psi}{2\mu L}; \quad r^2 = \frac{L - M - \psi}{\mu L}. \tag{16}$$

Introducing this into the first equation (14) we find

$$L = -\frac{\lambda}{2\mu L}\int_{L-M}^{L-M-\mu LR^2} F(\psi)\,d\psi. \tag{17}$$

The second equation (14) together with equation (17) now gives

$$M = L - M + \frac{\lambda}{2\mu L^2}\int_{L-M}^{L-M-\mu LR^2} \psi F(\psi)\,d\psi. \tag{18}$$

As has been said, the general properties of social imitation do not depend on the particular form of $F(\psi)$, as long as it satisfies the conditions mentioned above. We shall choose here, as an illustration

$$F(\psi) = \frac{\bar{N}_0 \gamma \psi}{1 + \gamma\psi}, \quad \text{for } \psi > 0, \tag{19}$$

where γ is a constant.

Since $F(\psi)$ is thus defined only for positive values of ψ, therefore the integrals in (17) and (18) have a meaning only if both limits are positive. This implies, in particular, $L - M > 0$. A priori we cannot be sure of this since L and M are to be determined by equations (17) and (18). We shall, however, treat those equations on the assumption that $L - M > 0$, and $L - M - L\mu R^2 > 0$, and then show that this leads actually to positive values of those two expressions, thus proving that such positive solutions exist.

Introducing (19) into (17) and evaluating the integral, we find

$$L = \bar{N}_0\left[\frac{\lambda R^2}{2} + \frac{\lambda}{2\mu L\gamma}\log\left(1 - \frac{\gamma LR^2\mu}{1 + \gamma(L - M)}\right)\right]. \tag{20}$$

In a similar way we find from (18)

$$M = L - M + \bar{N}_0\left[\frac{\lambda R^2}{2L}\frac{1 - \gamma(L - M)}{\gamma} + \frac{\lambda R^4\mu}{4} + \frac{\lambda}{2\mu L^2\gamma^2}\log\left(1 - \frac{\gamma LR^2\mu}{1 + \gamma(L - M)}\right)\right]. \tag{21}$$

The above expressions simplify considerably if μ is so small that

we may expand the log terms and preserve only the lowest non-vanishing powers of μ. This is quite in line with condition (8). With linear terms in μ only we find from (20)

$$L = \frac{\lambda \bar{N}_0 R^2}{2} \frac{\gamma(L-M)}{1+\gamma(L-M)}. \tag{22}$$

If we do the same thing with (21), we find $M = 0$, M being a difference of two finite terms. We must therefore preserve terms in μ^2 also. This gives

$$M = \frac{\lambda \bar{N}_0 R^4 \mu}{4} \frac{[1+\gamma(L-M)]^2 - 1}{[1+\gamma(L-M)]^2}. \tag{23}$$

Subtracting equation (23) from (22) and introducing a new variable

$$C = L - M, \tag{24}$$

we find

$$C = \frac{\lambda \bar{N}_0 R^2}{2} \frac{\gamma C}{1+\gamma C} - \frac{\lambda \bar{N}_0 R^4 \mu}{4} \left[1 - \frac{1}{(1+\gamma C)^2} \right]. \tag{25}$$

If μ is very small, so that the second term is negligible, equation (25) has the roots $C = 0$ and

$$C = \frac{\lambda \bar{N}_0 R^2}{2} - \frac{1}{\gamma}, \tag{26}$$

which is positive if

$$\lambda \gamma \bar{N}_0 R^2 > 2. \tag{27}$$

Equations (22) and (23) show that for $C > 0$, $L > 0$, and for a very small μ, $L - M - \mu L R^2 > 0$. If μ is not negligible, but small, so that

$$R^2 \mu << 2, \tag{28}$$

which almost follows from (8), equation (25) still has a positive root if (27) is satisfied, as is readily seen from graphing the right side of (25). If, however, μ becomes sufficiently large, so that the two terms of the right side of (25) are comparable, then equation (25) has only the root $C = 0$. Equations (22) and (23) then give $L = M = 0$, and from (15) it follows that $\psi = 0$.

We thus see that with a proper choice of constants, the integral equation (12) has a positive solution of the form (15). The adopted

behavior, in this case R_1, is exhibited by more individuals in the center of the area than at the periphery.

This may have definite sociological implications, in regard to behavior patterns characteristic for "the heart of a country." The physical reason for such a distribution of behavior is that even for uniform population density, an individual has more neighbors in the center than at the periphery. At the boundary he is surrounded by neighbors only on one side. Differences in behavior patterns in the heart of a country and at its borders are frequently due to the effects of adjacent countries with different behavior patterns. But the possible existence of the intrinsic effect discussed in this paper must be kept in mind.

It would be of interest to study areas which have the shape of either rectangles or ellipses, and see what effect the increase of the "specific boundary" would have. Better approximations than the inverted parabola should also be sought. If the distance effect on imitation is rather pronounced (rather large μ), then only neighbors in a limited region will have an effect. We still could use an inverted parabola, but extend the integration with respect to x',y' in equation (5) only to a circle of radius $1/\sqrt{\mu}$ around the point x,y. Inside of that radius $1 - \mu r_{12}^2 > 0$. Outside it is negative and we may interpret this as meaning that there is no influence from outside that circle. In this case the integral in (5) becomes a function of (x,y) not only by virtue of ψ containing those two variables, but also because x and y enter into the limits of integration. In such a case the results may possibly not depend too much on the size of the area S and a possibility of several maxima or minima for ψ should be investigated.

The restriction to sufficiently small values of μ seems to be connected with the circumstance that $K(r_{12})$ becomes negative for sufficiently large values of r_{12}. This is evidenced by the following consideration.

Put:

$$K_1(r,r') = 2\pi r' e^{-\mu(r^2+r'^2)}. \tag{29}$$

Introducing this into equation (6) and putting

$$Q = \lambda \int_0^R F[\psi(r)] e^{-\mu r^2} r\, dr, \tag{30}$$

we find

$$\psi(r) = Q e^{-\mu r^2}. \tag{31}$$

From (31) we have

$$r\,dr = -\frac{d\psi}{2\mu\psi}. \tag{32}$$

Remembering that for $R = 0$, $\psi = Q$, and for $r = R$, $\psi = \psi_0$, where ψ_0 is a very small quantity when R and μ are sufficiently large, we find from (30):

$$Q^2 = \frac{\lambda}{2\mu} \int_{\psi_0}^{Q} F(\psi)\,d\psi. \tag{33}$$

If $F(\psi)$ is of the general shape specified above, then its integral curve is symmetric with respect to $\psi = 0$, has a zero tangent at $\psi = 0$, is everywhere positive, and has two asymptotes, symmetric with respect to the line $\psi = 0$. Hence for a proper choice of λ/μ, regardless of the value of μ, equation (33) has two symmetric roots, one positive, the other negative.

It is, however, difficult to say to what expression $K(r_{12})$ corresponds the expression (29) for $K_1(r,r')$.

Essentially similar methods may be applied for the case in which $K(x,y;\ x',y')$ is not symmetric. This is likely to occur in many sociological situations. For example, an individual living in the city is less likely to imitate one living in the country than vice versa. The absolute geographical coordinates of the individual may thus affect his tendency to imitate or his property to be imitated. A convenient kernel $K(r,r')$ in this case may be

$$e^{-k(r-r')}$$

which will permit the use of the same method as has been used here.

We may generalize further the problem by considering a prescribed distribution of density $\overline{N}_0$.

Similar types of problems arise when we consider the effect of social distance on imitation. Nonsymmetrical functions of the type of (29) will be the rule here. The problem is somewhat simplified by being inherently one-dimensional.

LITERATURE

Landahl, H. D. 1950. "Mathematical Theory of Imitative Behavior in a Social Group with Finite Imitation Thresholds." *Bull. Math. Biophysics,* **12**, 208–213.

Landau, H. G. 1950. "Note on the Effect of Imitation in Social Behavior." *Bull. Math. Biophysics,* **12**, 221–236.

Rashevsky, N. 1949. "Mathematical Biology of Social Behavior: III." *Bull. Math. Biophysics,* **11**, 255–271.

——— 1950. *Mathematical Biology of Social Behavior.* Chicago: University of Chicago Press.

BULLETIN OF
MATHEMATICAL BIOPHYSICS
VOLUME 12, 1950

CONTRIBUTION TO THE PROBABILISTIC THEORY OF NEURAL NETS: II.

FACILITATION AND THRESHOLD PHENOMENA

ANATOL RAPOPORT
COMMITTEE ON MATHEMATICAL BIOLOGY
THE UNIVERSITY OF CHICAGO

The output curve of a single neuron with a threshold of response with respect to the frequency of the stimuli is derived. If the stimuli are regularly spaced in time, the output curve has discontinuities. If the threshold and/or refractory period are sufficiently large, the output curve approaches the "all-or-none" curve.

In the case of completely randomized stimuli, the output curve is sigmoid. The equation of this curve is derived and some properties are studied. Threshold and "all-or-none" effects can be achieved by "pyramiding" neurons of this type to converge on neurons of higher order.

In our discussion of the filter net (Rapoport, 1950a), we postulated a neuron possessing a threshold of response with respect to the frequency of incoming stimuli. The concept of threshold is fundamental in the theory of the propagation of nervous impulses. However, this concept varies in the several approaches to neural physiology. In all such discussions an "intensity" of the incoming stimulus plays a part. However, the question "intensity of what?" is answered differently by different authors and sometimes is not answered at all. Indeed, for the purposes of a quantitative theory it is often of advantage to omit all reference to the physical nature of the quantities considered, as has been done by Rashevsky, Hill, and others.

To be sure, in some models of the nervous system the specific character of "intensity" is to some extent determined by the model itself. Thus in the model of W. S. McCulloch and W. Pitts (1943) the "intensity" of a stimulus impinging on a cell body is implicitly defined as the number of end bulbs terminating upon the cell body which fire *simultaneously*. Thus in this model the threshold becomes simply the number of such bulbs which, when firing simultaneously, cause the neuron on which they impinge to fire. Here threshold is dimensionless. On the other hand, in our treatment (Rapoport,

1950a), intensity of input for a given neuron is defined as the *frequency* of stimuli received. Here threshold is a frequency and has dimensions $[T]^{-1}$.

The so called "all-or-none" law has likewise different interpretations in different models of the nervous system.

In the McCulloch-Pitts' picture the concept of "intensity of firing" is not applicable to a single neuron. The neuron either fires or does not fire, so that the all-or-none law applies to all neurons in that model. In our model the situation is different. Intensity (frequency) does have a meaning in connection with a single neuron. An all-or-none neuron (or aggregate), according to the frequency concept of intensity, would have to be one which responds with the same frequency (or not at all) to all stimulus frequencies. Such behavior is not an inherent characteristic of the model. If we wish to postulate neurons or aggregates which do exhibit an "all-or-none" behavior in this sense, we must derive conditions for this to be the case. In our construction of neural nets which simulate certain aspects of neurophysiological and psychological behavior we shall postulate the existence of such aggregates. Therefore, our problem is to derive the structure and parameters of such aggregates.

Regularly Spaced Stimuli.

The input-output curve of an "all-or-none" neuron is given by the following discontinuous function

$$\begin{aligned} f(x) &= 0, \quad \text{for} \quad x < x^*, \\ f(x) &= \text{constant} > 0, \quad \text{for} \quad x \geqslant x^*. \end{aligned} \tag{1}$$

The graph is shown in Figure 1.

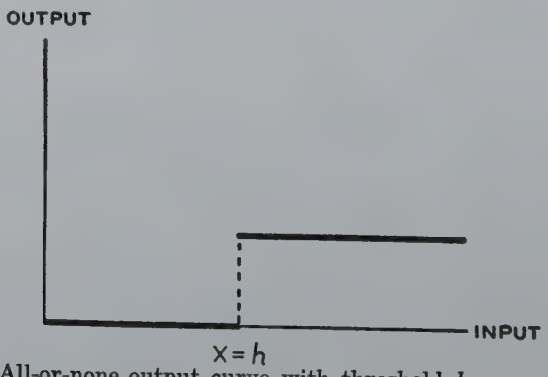

FIGURE 1. All-or-none output curve with threshold h.

If a neuron receives *regularly spaced* stimuli and has a certain threshold of response, its output curve will obviously contain a discontinuity at the threshold frequency of the input. In fact, let δ be the refractory period of the neuron and h the threshold. Then, if x is the input frequency, we have for the output

$$\begin{aligned} f(x) &= 0, \quad \text{for} \quad x < h, \\ f(x) &= x(h + [\delta x])^{-1}, \quad \text{for} \quad x \geqq h, \end{aligned} \tag{2}$$

where $[\delta x]$ is the greatest integer less than δx. The function (2) is discontinuous not only at $x = h$, but also at all the successive values of x for which δx is an integer, i.e., for $x = n/\delta$ where n takes the integral values, n_0, $n_0 + 1$, etc., and where n_0 is defined by the inequality

$$n_0 - 1 \leqq h\,\delta < n_0. \tag{3}$$

The discontinuities of $f(x)$ are thus equally spaced. The slope of $f(x)$ is likewise a discontinuous function

$$f'(x) = (h + [\delta x])^{-1}, \quad x \geqslant h, \tag{4}$$

where $[\delta x] = n_0, n_0 + 1$, etc.

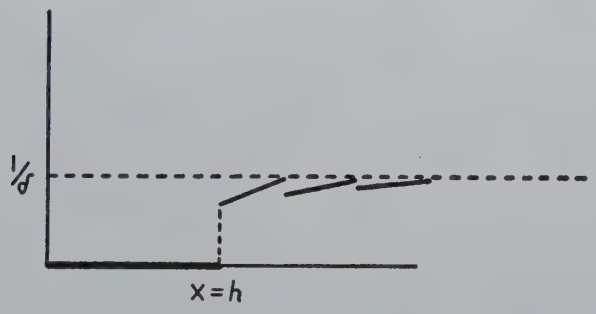

FIGURE 2. Input-output curve of a single neuron with a threshold.

The graph of (2) is shown in Figure 2. The approximation to the output of an all-or-none neuron (Figure 1) will be the closer, the smaller the slope $f'(x)$ immediately to the right of the first discontinuity. But this is

$$(h + n_0)^{-1} \sim [h(1 + \delta)]^{-1}. \tag{5}$$

Thus the quantity (5) can be said to measure the deviation from the all-or-none law.

Theorem 1. *If a neuron receives regularly spaced stimuli of which* h *must impinge per unit time to fire the neuron, the deviation*

from the all-or-none law is given by $[h(1 + \delta)]^{-1}$, *where* δ *is the refractory period of the neuron.*

Corollary. The all-or-none law is approximated the more closely the greater the threshold and the refractory period of the neuron.

Randomized Stimuli.

The foregoing case is probably of very limited physiological importance. If it is supposed that the incoming stimuli are the firings of a single neuron, then the threshold concept becomes dependent on "temporal summation," a hypothesis largely discredited among neurophysiologists. Indeed, it is impossible to account for temporal summation if the refractory periods exceed the periods of facilitation (or latent addition), because one would need to suppose that the neuron, whose firings are "summed," fires with a frequency greater than that allowed by its refractory period.

A more plausible assumption is that the stimuli come from several different neurons. Facilitation then depends on "spatial summation," a hypothesis generally accepted in neurophysiology. But then the difficulty arises in supposing the stimuli equally spaced. In the paper mentioned above of McCulloch and Pitts and in the subsequent work of N. Rashevsky (1946) and J. B. Roberts (1948) this assumption is fundamental. However, we consider the clock-work precision of synchronized firing a rather artificial hypothesis and will abandon it, even though it accounts very simply for frequency thresholds. We shall suppose instead, as in our previous paper, a complete randomization of incident stimuli and compute the input-output curve for a neuron with a facilitation mechanism (i.e., spatial summation of stimuli falling within a period of latent addition). We shall further inquire into the conditions under which the input-output curve may approach that of an all-or-none neuron or at least exhibit a threshold effect.

Consider a neuron receiving a shower of completely randomized stimuli of average frequency x per second from the outside and responding only to the hth stimulus of a group which falls within the interval σ, provided none of the h stimuli impinges during the refractory period.

Take the origin of time at δ seconds after the neuron has fired. Consider now the set of points on the time axis at which stimuli are received. Somewhere on the time axis there will be a first stimulus of a group of h stimuli which falls within σ. Call this first stimulus

s, and let us compute the probability distribution of the time of occurrence of s.

Since the incident stimuli are completely randomized, the probability of $h - 1$ additional stimuli occurring within σ seconds after s is independent of time, i.e., is a function of x and σ only. Call this function $g(x,\sigma)$. Then the probability of s falling in an arbitrarily chosen infinitesimal time interval dt will be

$$P(t)\,dt = \left[\, 1 - \int_0^t P(t)\,dt \,\right] (x\,dt)\, g(x,\sigma). \tag{6}$$

The right side of (6) is a product of three mutually independent probabilities, namely, in the order of factors: 1) that there has been no s in the interval $(0, t)$; 2) that a stimulus is impinging in dt; and 3) that $h - 1$ additional stimuli are impinging within σ seconds after s.

Dividing both sides by dt and differentiating with respect to t, we obtain the differential equation

$$P'(t) = -\,x g(x,\sigma) P(t), \tag{7}$$

which upon integration yields

$$P(t) = A \exp\{-\,x g(x,\sigma)\, t\}, \tag{8}$$

where A is the constant of integration.

Normalization of $P(t)$ demands that

$$A = x g(x,\sigma). \tag{9}$$

It remains to determine $g(x,\sigma)$.

Note that

$$g(x,\sigma) = \int_0^\sigma P^{(h-1)}(x,\tau)\,d\tau \tag{10}$$

where $P^{(h-1)}(x,t)$ is the probability distribution of the time of the $(h - 1)$th event of a Poisson-distributed sequence of average frequency x. The recurrence formula for these distributions is given by (Jost, 1947)

$$P^{(k)}(t) = \int_0^t P^{(k-1)}(t-\tau) P(\tau)\,d\tau, \tag{11}$$

where $P(\tau)$ is the distribution of the first event. In the case of the Poisson distribution, this leads to

$$P^{(h-1)}(t) = \frac{x^{h-1}}{(h-2)!}\, t^{h-2} e^{-xt}. \tag{12}$$

Thus

$$g(x,\sigma) = \int_0^\sigma \frac{x^{h-1}}{(h-2)!}\,\tau^{h-2}\,e^{-x\tau}\,d\tau. \tag{13}$$

We define the polynomials $E_k(z)$ as partial expansions of exp $\{z\}$, broken off at the kth power of the argument. Thus

$$E_k(z) \equiv \sum_{j=0}^{k} z^j/j!. \tag{14}$$

Integrating the right side of (13) by parts and simplifying gives

$$g(x,\sigma) = 1 - E_{h-2}(x\sigma)\,e^{-x\sigma}. \tag{15}$$

But equation (8) shows that the distribution of the instant s is a Poisson distribution with average frequency $xg(x,\sigma)$. Therefore the expected time of occurrence of s, counted from any instant δ or more seconds following a firing instant, is given by

$$\{x[1 - E_{h-2}(x\sigma)\,e^{-x\sigma}]\}^{-1}. \tag{16}$$

Our next step is to compute the expected time of occurrence of the firing, following the occurrence of s. This is clearly given by averaging the time of occurrence of the $(h-1)$th stimulus following s over the interval σ. This average is given by

$$\begin{aligned}
&\frac{\displaystyle\int_0^\sigma \frac{x^{h-1}\,\tau^{h-1}}{(h-2)!}\,e^{-x\tau}\,d\tau}{\displaystyle x\int_0^\sigma \frac{x^{h-2}\,\tau^{h-2}}{(h-2)!}\,e^{-x\tau}\,d\tau} \\
&= \frac{e^{-x\sigma}[(x\sigma)^{h-1} + (h-1)(x\sigma)^{h-2} + \cdots (h-1)!] - (h-1)!}{xe^{-x\sigma}[(x\sigma)^{h-2} + (h-2)(x\sigma)^{h-3} + \cdots (h-2)!] - (h-2)!} \\
&= \frac{(h-1)[1 - e^{-x\sigma}E_{h-1}(x\sigma)]}{x[1 - e^{-x\sigma}E_{h-2}(x\sigma)]}.
\end{aligned} \tag{17}$$

For very large σ or for very large x, the expected time of the $(h-1)$th incidence, measured from *any* incident stimulus, falls within σ and is approximately equal to $(h-1)/x$, as is seen from the last expression of equation (17).

We can now write down the total expected time $\bar{t}$ of the next

firing from an instant of firing of our neuron. This will be seen to be the sum of 1) the refractory period; 2) the expected time of s; and 3) the expected time of the $(h-1)$th stimulus following s. Hence

$$\bar{t} = \delta + \frac{1 + (h-1)[1 - e^{-x\sigma} E_{h-1}(x\sigma)]}{x[1 - e^{-x\sigma} E_{h-2}(x\sigma)]}. \tag{18}$$

Therefore, the average frequency, given by the reciprocal of $\bar{t}$ will be

$$f(x, \delta, \sigma, h) = \frac{x[1 - e^{-x\sigma} E_{h-2}(x\sigma)]}{\delta x[1 - e^{-x\sigma} E_{h-2}(x\sigma)] + 1 + (h-1)[1 - e^{-x\sigma} E_{h-1}(x\sigma)]}. \tag{19}$$

The Properties of the Facilitation Input-Output Curve.

We wish to examine the behavior of the function (19) for small and large values of x, where the other parameters are held fixed. To do this, we first examine the expression $[1 - e^{-x\sigma} E_{h-2}(x\sigma)]$. This can be written as

$$e^{-x\sigma}[e^{x\sigma} - E_{h-2}(x\sigma)] = e^{-x\sigma} \sum_{k=(h-1)}^{\infty} \frac{(x\sigma)^k}{k!} = \sum_{k=0}^{\infty} \frac{(-x\sigma)^k}{k!} \cdot \sum_{k=h-1}^{\infty} \frac{(x\sigma)^k}{k!}. \tag{20}$$

This product of two infinite series is seen to be itself an infinite series whose lowest power term is $(x\sigma)^{h-1}/(h-1)!$. But then the numerator of the right side of (19) is also a series, whose lowest power term is $x^h\sigma^{h-1}/(h-1)!$ and this is the dominant term for small values of x. On the other hand, the denominator of the right side of (19) approaches unity as x becomes vanishingly small. We thus have

Theorem 2. *The facilitation input-output curve for randomized stimuli behaves like* $x^h\sigma^{h-1}/(h-1)!$ *for small values of* x.

For very large x, the first term of the denominator of (19) becomes dominant so that the other two may be neglected. But then, canceling the brackets in the numerator and denominator, we obtain $1/\delta$ for the average frequency, which should be the case.

Let us now consider the behavior of the function (19) with respect to some limiting values of h. If we define $E_{-1}(z) \equiv 0$, which is consistent with the general definition of $E_k(z)$, since E_k may be defined by the recurrence relation

$$\frac{dE_k(z)}{dz} = E_{k-1}(z); \qquad E_k(0) = 1, \tag{21}$$

then it is easily seen that for $h = 1$, the expression (19) reduces to

$$f(x,\delta) = x(\delta x + 1)^{-1}. \tag{22}$$

This is the output of a single neuron responding to every stimulus of a Poisson shower, as previously shown (Rapoport, 1950a). On the other hand, if h is large, then, since,

$$\lim_{k\to\infty} [E_{k+1}(z) - E_k(z)] = \lim_{k\to\infty} \frac{z^k}{k!} = 0, \tag{23}$$

we may write with good approximation for (19)

$$f(x, \delta, \sigma) = \frac{xq(x\sigma)}{(\delta x + h - 1)q(x\sigma) + 1}, \tag{24}$$

where $q(x\sigma) \equiv 1 - e^{-x\sigma} E_{h-2}(x\sigma)$. Formula (24) indicates how the simple output curve (22) is modified by the introduction of facilitation together with a large threshold.

The graph of (19) is shown in Figure 3. It is a sigmoid curve

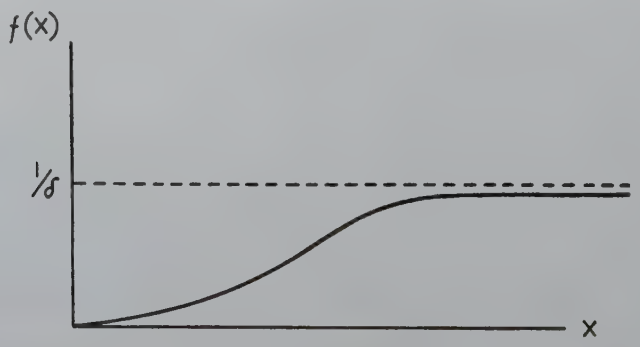

FIGURE 3. Output of a neuron with facilitation, period receiving stimuli randomized in time.

depending on the parameters δ, σ, and h, of which only two are essential, since the choice of time units may make either δ or σ equal to one. To fix ideas, let $\delta = 1$. The question arises whether the parameters σ and h can be so chosen that the curve approaches the all-or-none curve, or at least exhibits a threshold effect as in Figure 2. That is to say, can σ and h be so chosen that the curve starts out with negligible slope for a finite range of the input and suddenly rises to almost its maximum value, i.e., the slope in the neighbor-

hood of the inflexion point becomes very large, while it is small everywhere else? This question may be of importance in the mathematical theory of evolving organisms in the following way. Suppose that an organism has mechanisms which depend on neural aggregates responding according to an all-or-none law or having a frequency threshold of response. If it could be shown that such mechanisms can be constructed of neurons such as described in this paper, one could make a hypothesis that such neurons arose through natural selection, where those individuals were successively selected whose parameters σ and h approached gradually their "optimum" values.

Our conjecture is that such is *not* the case, because there is a finite upper bound for the values of $\partial f(x, \delta, \sigma, h)/\partial x$ whatever be the values of σ and h. The justification of the conjecture will be investigated elsewhere. We will, however, present a different mechanism based on neurons of the type described here which does exhibit threshold and all-or-none effects.

Threshold Effect by Pyramiding.

Consider a net as shown in Figure 4 consisting of an aggregate of n neurons N_1, each of which has an output curve as in Figure 3 and all of which converge on N_2, a neuron of higher order. For simplicity all the parameters δ, σ, and h are taken to be the same for all neurons. If N is large, the firings of the neurons N_1 will be sufficient-

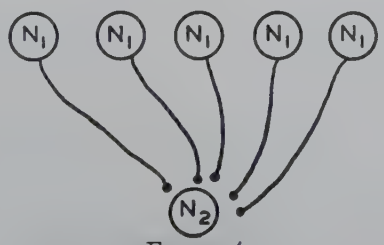

FIGURE 4

ly randomized so that the same mathematical treatment is applicable to the output curve of N_2 as has been given above, except that the input will be the combined total output of the neurons N_1. We will then have for the output of N_2

$$F(x) = f[nf(x)]. \tag{25}$$

The above discussion of $f(x)$ has shown that it is a sigmoid curve with zero initial slope and a constant limiting value. To see how such a curve is modified by the transformation (25), we must

compare the effects of the inputs x and $nf(x)$. Figure 5 shows a simultaneous graph of the functions $y = x$ and $y = nf(x)$. Note that $f(x)$ is a monotone increasing function. Thus there always exists a

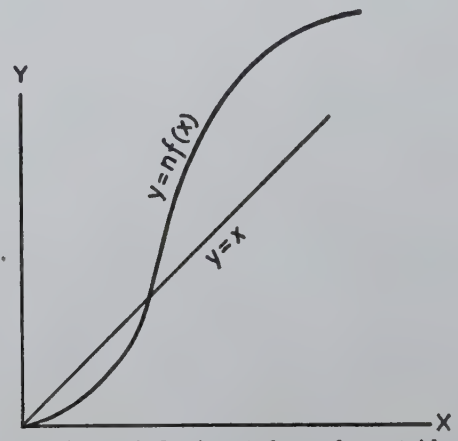

FIGURE 5. A comparison of the input from the outside, $y = x$ with the input from n neurons N_1, $y = nf(x)$.

value $x = x^*$, such that $nf(x) < x$ for $x < x^*$ and $nf(x) > x$ for $x > x^*$. If we simultaneously plot $f(x)$ and $F(x)$, we will also have $F(x) < f(x)$ for $x < x^*$ and $F(x) > f(x)$ for $x > x^*$ as shown in Figure 6.

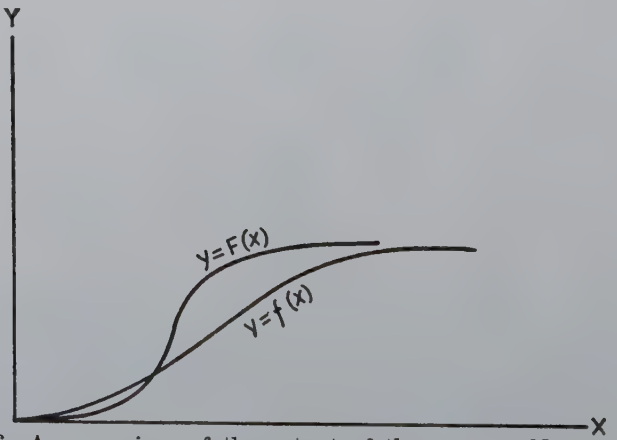

FIGURE 6. A comparison of the output of the neurons N_1, $y = f(x)$ with the output of the neuron N_2, $y = F(x) = f[nf(x)]$.

Since n can be taken arbitrarily large, the effectiveness of depressing and raising $F(x)$ on both sides of x^* can be made arbitrarily large. It is true, however, that x^* is a function of n, in fact,

is vanishingly small when n is very large. This spoils the main feature of the threshold effect, since it depresses the threshold to zero. However, this situation can be remedied by increasing h. A large h flattens $f(x, h)$ and, therefore, $nf(x, h)$ near the origin. Therefore a large h has an effect opposite to that of large n. Hence we see that by choosing n and h sufficiently large, we can approach the all-or-none effect with respect to frequency of stimuli by a pyramiding arrangement.

This investigation is part of the work done under Contract No. AF19(122)-161 between the U. S. Air Force Cambridge Research Laboratories and The University of Chicago.

LITERATURE

Jost, Res. 1947. "Bemerhungen zur mathematischen theorie der zähler." *Helvetica Physica Acta*, **20**, 173.

McCulloch, W. S. and W. Pitts. 1943. "A Logical Calculus of the Ideas Immanent in Nervous Activity." *Bull. Math. Biophysics*, **5**, 115–133.

Rapoport, A. 1950a. "Contribution to the Probabilistic Theory of Neural Nets: I." *Bull. Math. Biophysics*, **12**, 109–121.

Rashevsky, N. 1946. "A Neural Mechanism for Logical Thinking." *Bull. Math. Biophysics*, **8**, 29–40.

Roberts, J. B. 1948. "Interaction of Cycles." *Bull. Math. Biophysics*, **10**, 123–129.

MATHEMATICAL BIOPHYSICS
VOLUME 12, 1950

ON A MATHEMATICAL THEORY OF THE REACTION OF CELLS TO X-RAY IRRADIATION

EDMUND PINNEY
UNIVERSITY OF CALIFORNIA

To describe the variation of concentration of certain substances in cells subjected to X-rays for a finite length of time, R. M. Sievert (1941) has given an equation of mixed differences involving time-lag and having discontinuous coefficients. In the present paper Laplace transform methods are applied to solve this equation in terms of a series valid for all values of time. In particular, the conditions given by J. Th. van der Werff (1943) that the concentration approaches a constant value are confirmed; otherwise steady oscillations in concentration may be expected, or oscillations appear whose amplitude steadily increases until the cell is destroyed or the assumed mathematical model otherwise fails to apply.

R. M. Sievert (1941) has studied the concentration of certain substances on cells under the action of X-ray irradiation. Assuming the reaction of the cell to deviations from the equilibrium concentration to be reversible and to occur with a time-lag τ, Sievert obtained

$$\frac{dx(t)}{dt} = -I\alpha_t x(t) + R\beta_t[x_0 - x(t-\tau)], \tag{1}$$

where $x(t)$ measures the concentration of the substance, x_0 is an equilibrium value of $x(t)$, I is an irradiation constant depending upon the incoming radiation, and R is a constant measuring the reaction of the cell to a departure from the equilibrium concentration. The quantities α_t and β_t are defined by

$$\begin{aligned} \alpha_t &= 1, \quad 0 < t < T, \\ &= 0, \quad t > T; \end{aligned} \tag{2}$$

$$\begin{aligned} \beta_t &= 0, \quad 0 < t < \tau, \\ &= 1, \quad t > \tau, \end{aligned} \tag{3}$$

T being the irradiation time. The irradiation constant I is positive or negative according to whether the effect of the radiation is to decrease or increase the concentration of the substance.

Sievert has given graphical solutions for some particular values of the parameters. J. Th. van der Werff (1943) has applied the method of generating functions to Sievert's equations and determined the conditions under which the solutions asymptotically approach constant values. He has shown that, as $t \to \infty$,

$$\begin{aligned} &x(t) \to x_0 \text{ when } 0 \leqslant T < \infty \text{ and } R_\tau < \pi/2, \\ &x(t) \to x_0 R/(R+I) \text{ when } T = \infty \text{ and } R \leqslant I, \\ &x(t) \to x_0 R/(R+I) \text{ when } T = \infty, R > I, R_\tau < \phi/\sin\phi, \end{aligned} \tag{4}$$

where ϕ is the smallest positive root of $\cos\phi + I/R = 0$.

Other types of asymptotic behavior (oscillations, for instance) might be physically conceivable. More powerful methods are now available, so the asymptotic behavior may be determined for all ranges of the parameters. A solution for $x(t)$, valid for all positive t, will be given in this paper, and this will, incidentally, give the asymptotic behavior desired.

It follows from two theorems of E. M. Wright (1948, Theorems 1 and 3) that a unique differentiable solution $x(t)$ of (1) exists which, together with its derivative, is exponentially bounded as $t \to \infty$, that is, real constants K, s_0 exist such that as $t \to \infty$

$$|x(t)| < Ke^{s_0 t}, \quad |x'(t)| < Ke^{s_0 t}. \tag{5}$$

Therefore the Laplace transform of $x(t)$ and $x'(t)$ exists for $Re(s) > s_0$. We denote the Laplace transform by

$$X(s) = \int_0^\infty e^{-st} x(t)\, dt, \quad Re(s) > s_0. \tag{6}$$

Integrating by parts, taking the initial value of $x(t)$ as the equilibrium value x_0, there results

$$sX(s) - x_0 = \int_0^\infty e^{-st} x'(t)\, dt, \quad Re(s) > s_0. \tag{7}$$

Multiply equation (1) by e^{-st} and integrate with respect to t from 0 to ∞. Using (2), (3) and (7), we obtain

$$sX(s) - x_0 = Rx_0 \int_\tau^\infty e^{-st} dt - x_0 ITF(sT) - R \int_\tau^\infty e^{-st} x(t-\tau)\, dt,$$

where

$$F(sT) = \frac{1}{x_0 T} \int_0^T e^{-st} x(t)\, dt. \tag{8}$$

Replacing t by $t + \tau$ in the last integral of the above expression and using (6) we have

$$(s + Re^{-s\tau})X(s) = x_0\left[1 + \frac{R}{s}e^{-s\tau} - ITF(sT)\right]. \tag{9}$$

Using the inverse Laplace transformation (Doetsch, 1943; Satz 6[6.5], p. 107), we find

$$x(t) = \frac{1}{2\pi i}\int_{p-i\infty}^{p+i\infty} X(s)e^{st}ds \tag{10}$$

where $p > s_0$. That is, all the poles of $X(s)$ lie to the left of the line $Re(s) = p$.

In calculating $x(t)$ from (9) and (10), it is necessary to know something about the roots of the transcendental equation

$$\sigma e^{\sigma} + r = 0 \tag{11}$$

in the case $r > 0$. This equation has been studied by a number of authors, the first probably being L. Euler (1777). The work of F. Schürer (1912, p. 175) is one of the most complete treatments of this type. When $0 < r < 1/e$, (11) has two distinct negative real roots which we will denote by $\sigma_{0+}(r)$ and $\sigma_{0-}(r)$, where $\sigma_{0-}(r) < \sigma_{0+}(r) < 0$. As r increases to $1/e$, these roots coalesce to a double root, $\sigma_{0\pm}(1/e) = -1$. As r increases above $1/e$, these real roots disappear altogether; in their place a pair of conjugate complex roots appears. This pair will also be denoted by $\sigma_{0\pm}(r)$, the plus sign pertaining to the root with positive imaginary part, and the minus sign pertaining to the root with negative imaginary part.

In addition, (11) has infinitely many other pairs of conjugate complex roots. Schürer has shown that these roots all exceed the roots $\sigma_{0\pm}(r)$ in absolute magnitude and in fact may be ordered. $\sigma_{1\pm}(r)$, $\sigma_{2\pm}(r)$, $\cdots$ in such a way that

$$|\sigma_{0+}(r)| \leqslant |\sigma_{0-}(r)| < |\sigma_{1\pm}(r)| < |\sigma_{2\pm}(r)| < \cdots. \tag{12}$$

Setting

$$\sigma = \alpha + i\beta \tag{13}$$

in (11) and equating real and imaginary parts to zero, we have

$$e^{\alpha}(\alpha\cos\beta - \beta\sin\beta) + r = 0, \quad \beta\cos\beta + \alpha\sin\beta = 0.$$

Solving the second equation for α and inserting this in the first we obtain

$$\alpha = -\beta/\tan\beta\,, \tag{14}$$

$$f(\beta) \equiv \beta/\tan\beta - \ln(\beta/\sin\beta) = -\ln r\,. \tag{15}$$

The quantity β may be determined from (15) and α from (14).

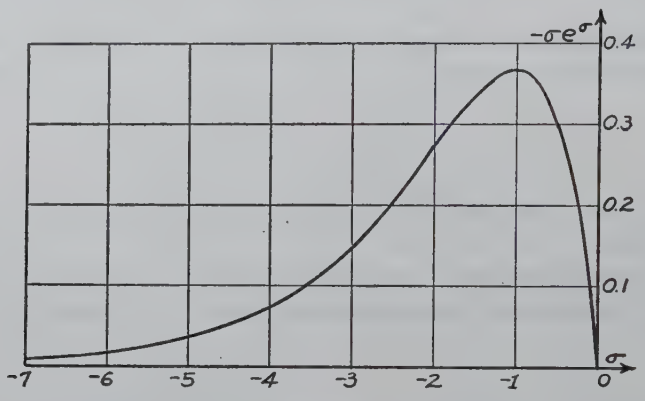

FIGURE 1

The real roots may be determined graphically from Figure 1 and equation (11). The complex roots may also be determined graphically. The function $f(\beta)$ has been tabulated by R. Frisch and H. Holme (1935) and is shown graphically in Figure 2. From this graph and (15), the successive values $\beta_0(r)$, $\beta_1(r)$, $\beta_2(r)$, $\cdots$ of β can be determined for each r. Then, using (14), the corresponding values $\alpha_0(r)$, $\alpha_1(r)$, $\alpha_2(r)$, $\cdots$ of α can be determined from

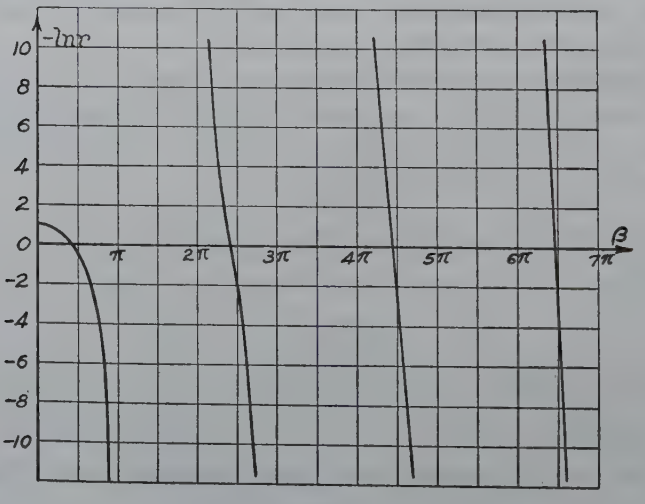

FIGURE 2

$$\alpha_n(r) = -\beta_n(r)\cot\beta_n(r). \tag{16}$$

The complex roots of (11) are then

$$\sigma_{n\pm}(r) = \alpha_n(r) \pm i\beta_n(r). \tag{17}$$

From equation (15) we see that β_n is a first or second quadrant angle. As $n \to \infty$, $\beta_n \to \infty$ for otherwise the roots of (15) would have a point of condensation, implying $f(\beta) \equiv 0$ in its region of analyticity about this point, which is clearly not the case. As $\beta \to \infty$, the first term of $f(\beta)$ in (15) dominates, and, in fact, approaches infinity unless $\beta \to (2n + 1/2)\pi$, i.e., $\beta_n = (2n + 1/2)\pi - \varepsilon_n$, where $\varepsilon_n \to 0$ as $n \to \infty$. Then $\sin\beta_n = 1 + 0(\varepsilon_n^2)$, $\cot\beta_n = \varepsilon_n + 0(\varepsilon_n^3)$, so (15) gives $(2n + 1/2)\pi\varepsilon_n - \ln[(2n + 1/2)\pi/r] = 0(n\varepsilon_n^3) + 0(\varepsilon_n^2) + 0(\varepsilon_n/n)$. Therefore $\varepsilon_n = \delta_n + 0(\varepsilon_n^3) + 0(\varepsilon_n^2/n) + 0(\varepsilon_n/n^2) = \delta_n + 0(\delta_n^3)$, where

$$\delta_n = [(2n + 1/2)\pi]^{-1}\ln[(2n + 1/2)\pi/r].$$

Using (14) and (15) we find

$$\begin{aligned}\alpha_n &= -\ln[\beta_n/(r\sin\beta_n)] = -\ln\{\sec\varepsilon_n[(2n + 1/2)\pi - \varepsilon_n]/r\}\\ &= -\ln[(2n + 1/2)\pi/r] - \varepsilon_n^2/2 + \varepsilon_n/[(2n + 1/2)\pi] + 0(\varepsilon_n^4).\end{aligned}$$

Therefore, for sufficiently large n we find

$$\begin{aligned}\alpha_n(r) &= -\ln[(2n + 1/2)\pi/r] - \delta_n^2/2\\ &\quad + \delta_n/[(2n + 1/2)\pi] + 0(\delta_n^4),\\ \beta_n(r) &= (2n + 1/2)\pi - \delta_n + 0(\delta_n^3),\\ \delta_n &= [(2n + 1/2)\pi]^{-1}\ln[(2n + 1/2)\pi/r].\end{aligned} \tag{18}$$

Transposing r to the right side of (11), dividing by σ, taking the absolute value, and applying (12) and (13) we obtain

$$|\alpha_{0+}(r)| \geqslant |\alpha_{0-}(r)| > |\alpha_1(r)| > \alpha_2(r) > \cdots. \tag{19}$$

Consider the special case in which $T = \infty$. Equations (6) and (8) imply that $TF(sT) \to X(s)/x_0$ as $T \to \infty$. Using (9),

$$X(s) = \frac{x_0}{s}\frac{s + Re^{-s\tau}}{s + I + Re^{-s\tau}}. \tag{20}$$

This has poles at $s = 0$ and at $s = -I + (1/\tau)\sigma_{n\pm}(R\tau e^{I\tau})$ for $n = 0, 1, 2, \cdots$. The function $X(s)$ is bounded on the circle at infinity. Therefore the contour in (10) may be completed by adding the left-hand semi-circle at infinity when $t > 0$. Then applying Cauchy's theorem, for $R\tau e^{I\tau} \neq 1/e$, we find

$$x(t) = \frac{x_0 R}{I+R} + x_0 I\tau \sum_{n=0}^{\infty} \left\{ \frac{\exp[-It + t\sigma_{n+}(R\tau e^{I\tau})/\tau]}{[I\tau - \sigma_{n+}(R\tau e^{I\tau})][1 + \sigma_{n+}(R\tau e^{I\tau})]} + \frac{\exp[-It + t\sigma_{n-}(R\tau e^{I\tau})/\tau]}{[I\tau - \sigma_{n-}(R\tau e^{I\tau})][1 + \sigma_{n-}(R\tau e^{I\tau})]} \right\}. \tag{21}$$

From (17) and (18) we see that these series are uniformly convergent in t for $t \geqslant 0$, and all but at most a finite number of terms of the series tend exponentially to zero at $t \to \infty$. It is not difficult to show that the second and third conditions of (4) are precisely the conditions that all the exponents in (21) have negative real parts, so the last two results of (4) are verified. If $R > I$, $R\tau = \phi/\sin\phi$, where ϕ is the smallest positive root of $\cos\phi + I/R = 0$, $x(t)$ oscillates with a frequency ϕ/τ and an amplitude which may be obtained by setting $\sigma_{0\pm}(R\tau e^{I\tau}) = -I\tau \pm i\phi$ in the $n = 0$ terms of (21). Finally, for $R\tau > \phi/\sin\phi$, $x(t)$ undergoes a negatively damped oscillation which increases in amplitude until the regime of equation (1) ceases to apply.

Now consider the case in which $T < \infty$. In this case the solution (21) applies in the interval $0 < t < T$. Therefore, using (8) and (21) we obtain

$$\begin{aligned} F(s) = {} & \frac{R}{I+R} \cdot \frac{1}{s}[1 - e^{-s}] \\ & + I\tau \sum_{m=0}^{\infty} \{1 - \exp[-s - IT + \sigma_{m+}(R\tau e^{I\tau})T/\tau]\} \\ & \times \{[s + IT - \sigma_{m+}(R\tau e^{I\tau})T/\tau][I\tau - \sigma_{m+}(R\tau e^{I\tau})] \\ & \times [1 + \sigma_{m+}(R\tau e^{I\tau})]\}^{-1} \\ & + I\tau \sum_{m=0}^{\infty} \{1 - \exp[-s - IT + \sigma_{m-}(R\tau e^{I\tau})T/\tau]\} \\ & \times \{[s + IT - \sigma_{m-}(R\tau e^{I\tau})T/\tau][I\tau - \sigma_{m-}(R\tau e^{I\tau})] \\ & \times [1 + \sigma_{m-}(R\tau e^{I\tau})]\}^{-1}. \end{aligned} \tag{22}$$

In the particular case when $T < \tau$, a simpler expression may be obtained. For $0 < t < T$, using (1), (2) and (3), we have

$$x(t) = x_0 e^{-It}. \tag{23}$$

Using (8) we also have

$$F(s) = \frac{1}{s + IT}\left[1 - e^{-(s+IT)}\right]. \tag{24}$$

The expressions on the right-hand sides of (22) and (24) are without singularities except on the right-hand semi-circle at infinity. If either expression is substituted in (10) and the contour of (10) is completed, as it may be for $t > T - \tau$ by adding the left-hand semi-circle at infinity, Cauchy's theorem may be applied, giving

$$x(t) = x_0\left[1 - IT\sum_{n=0}^{\infty}\left\{\frac{F(\sigma_{n+}(R_\tau)T/\tau)}{1+\sigma_{n+}(R_\tau)}\exp\left[t\sigma_{n+}(R_\tau)/\tau\right] + \frac{F(\sigma_{n-}(R_\tau)T/\tau)}{1+\sigma_{n-}(R_\tau)}\exp\left[t\sigma_{n-}(R_\tau)/\tau\right]\right\}\right], \tag{25}$$

where $F(s)$ is obtained from either (22) or (24).

From (25) we see that $x(t) \to x_0$, provided all the exponents on the right-hand side have negative real parts. Because of equation (19) this will be true if $\alpha_{0+}(R_\tau) < 0$. For $R_\tau < 1/e$, $\sigma_{0\pm}(R_\tau)$ are negative real roots, so this is true. For $R_\tau > 1/e$, $\sigma_{0\pm}(R_\tau)$ are conjugate complex. From (16), we have $\alpha_{0+}(R_\tau) < 0$ if $\beta_0(R_\tau)$ lies in the first quadrant. From (15) and Figure 2 we see that the condition for this is that $R_\tau < \pi/2$. The first part of (4) is thus verified.

When $R_\tau = \pi/2$, $\beta_0 = \pi/2$, $x(t)$ oscillates with a frequency $\pi/(2\tau) = R$ and an amplitude which may be obtained by setting $\sigma_{0\pm}(R_\tau) = \pm \pi i/2$ in the $n = 0$ terms of (25). Finally, for $R_\tau > \pi/2$, $x(t)$ undergoes a negatively damped oscillation which increases in amplitude until the regime of equation (1) ceases to apply.

Equation (1) can be expected to be only a more-or-less rough approximation to the actual behavior of a cell under X-ray irradiation. A better approximation might involve non-linear terms. Generally the effect of non-linearities is to spread out the region of undamped oscillations. Thus one might expect undamped oscillations in an interval about $R_\tau = \phi/\sin\phi$ in the case $T = \infty$, and in an interval about $R_\tau = \pi/2$ in the case $T < \infty$.

This work was done in connection with the Office of Naval Research contract N6-0NR 251 TO 2 at Stanford University.

LITERATURE

Doetsch, G. 1943. *Theorie und Anwendung der Laplace-Transformation.* Dover.

Euler, L. 1777. *Acta Ac. Sci. Petropolitanae,* I, Part 1, p. 37.

Frisch, R. and H. Holme. 1935. "The characteristic equations of a mixed difference and differential equation occurring in economic dynamics." *Econometrica,* 3, 225–239.

Schürer, F. 1912. "Über die Funktional-Differentialgleichung $f'(x + 1) = af(x)$." *Ber. Verh. Sächs Akad. Wiss. Leipzig. Math.-Nat. Kl.,* 64, 167–236.

Sievert, R. M. 1941. "Zur theoretische-mathematischer Behandlung des Problems der biologischen Strahlenwirken." *Acta Radiologica*, **22**, 237–251.

van der Werff, J. Th. 1943. "Over enkele functionaal vergelijkingen reactieverschijnselen na Rontgenbestraling." *Nieuw Archief voor Wiskunde*, (2) **21**, 197–211.

Wright, E. M. 1948. "Linear difference differential equations." *Proc. Cambridge Philosophical Soc.*, **44**, 179–185.

BULLETIN OF
MATHEMATICAL BIOPHYSICS
VOLUME 12, 1950

MATHEMATICAL THEORY OF IMITATIVE BEHAVIOR IN A SOCIAL GROUP WITH FINITE IMITATION THRESHOLDS

H. D. LANDAHL
COMMITTEE ON MATHEMATICAL BIOLOGY
THE UNIVERSITY OF CHICAGO

The effect of the introduction of either a constant or a variable threshold into Rashevsky's theory of imitative behavior is considered. It is found that in either case it is possible to have five equilibrium configurations, three of which may be stable.

A theory of imitative behavior has been developed by N. Rashevsky (1949) in which two mutually exclusive activities are considered. The following discussion presupposes the reader's familiarity with that paper. The tendency to perform either activity R_1 or R_2 is considered to be determined by the observed performance of others. This tendency, together with the inherent individual bias and random fluctuations in the discrimination mechanism (Landahl, 1938), determines the final response. Let $p(\Delta)$ be the distribution of fluctuations, which is approximately a normal error curve, and let $N(\phi)$ be the distribution of biases in favor of R_1. If ψ is a measure of the tendency to prefer R_1 to R_2 owing to imitation, and the intensity of excitation is proportional to the difference between the numbers X and Y exhibiting R_1 and R_2 respectively, then ψ is given by [Rashevsky, 1949, equation (8)]

$$\frac{d\psi}{dt} = A(X - Y) - a\psi. \tag{1}$$

Constant Threshold. Using the same arguments as those of N. Rashevsky (1949), but now introducing a threshold h (Landahl, 1938), we find for the number X of individuals exhibiting behavior R_1, the expression

$$X = \int_{-\infty}^{\infty} N(\phi) \int_{-\psi-\phi+h}^{\infty} p(\Delta)\, d\Delta\, d\phi\,. \tag{2}$$

The number Y which exhibits behavior R_2 is given by a similar expression, except that the limits of integration with respect to Δ are $-\infty$ and $(-\psi - \phi - h)$. The number showing neither R_1 nor R_2 then is $N_0 - X - Y$, where N_0 is the total number.

Using the notation

$$2N_0 I(\psi) = \int_{-\infty}^{\infty} N(\phi) \left[\int_{-\psi-\phi+h}^{\infty} p(\Delta)\, d\Delta - \int_{-\infty}^{-\psi-\phi-h} p(\Delta)\, d\Delta \right] d\phi, \quad (3)$$

we find that equation (1) may be written as

$$\frac{d\psi}{dt} = 2AN_0 I(\psi) - a\psi. \quad (4)$$

We shall now evaluate the function $I(\psi)$ for the case in which both $p(\Delta)$ and $N(\phi)$ are normal distribution functions, but only $p(\Delta)$ is necessarily symmetric with respect to $\Delta = 0$.

Let

$$g(\zeta) = \frac{1}{\sqrt{2\pi}} e^{-\zeta^2/2} \quad (5)$$

and

$$G(x) = \int_0^x g(\zeta)\, d\zeta. \quad (6)$$

If $x = \phi/\sigma$, σ being the standard deviation and $\phi_0 = \sigma x_0$ being the mean value of the function $N(\phi)$, then $N(\phi)\, d\phi = N_0 g(x - x_0)\, dx$. Thus if s is the standard deviation of $p(\Delta)$, equation (3) becomes

$$2I(\psi) = \int_{-\infty}^{\infty} g(x - x_0) \times \left[-G\left(\frac{-\phi - \psi + h}{s}\right) - G\left(\frac{-\phi - \psi - h}{s}\right) \right] dx. \quad (7)$$

We now introduce the notation

$$\beta = \psi/s, \quad \alpha = \sigma/s, \quad H = h/s. \quad (8)$$

If $I(\beta s)$ $[= I(\psi)]$ is differentiated with respect to β we obtain

$$\frac{\partial I(\beta s)}{\partial \beta} = \frac{1}{2\sqrt{1+\alpha^2}}\, g\left(\frac{\alpha x_0 + \beta + H}{\sqrt{1+\alpha^2}}\right) + \frac{1}{2\sqrt{1+\alpha^2}}\, g\left(\frac{\alpha x_0 + \beta - H}{\sqrt{1+\alpha^2}}\right). \tag{9}$$

Integrating this with respect to β from zero to β we obtain (cf. Landau, 1950)

$$2I(\beta s) = G\left(\frac{\alpha x_0 + H + \beta}{\sqrt{1+\alpha^2}}\right) + G\left(\frac{\alpha x_0 - H + \beta}{\sqrt{1+\alpha^2}}\right) - G\left(\frac{\alpha x_0 + H}{\sqrt{1+\alpha^2}}\right) - G\left(\frac{\alpha x_0 - H}{\sqrt{1+\alpha^2}}\right). \tag{10}$$

At equilibrium, $d\psi/dt = 0$ and, therefore, for equilibrium values ψ_e of ψ, equation (1) reduces to the following expression:

$$I(\psi_e) = a\psi_e/2AN_0. \tag{11}$$

If $I(\psi) > a\psi/2AN_0$, then, from equation (1), $d\psi/dt > 0$, and, hence, ψ increases. Conversely, ψ decreases if the inequality sign is reversed. Thus we may find out whether or not a point of equilibrium is stable by showing that a small displacement away from a particular value ψ_e results in a spontaneous return to that value (cf. Rashevsky, 1949).

We shall consider only the case in which $x_0 = 0$. If we introduce the notation

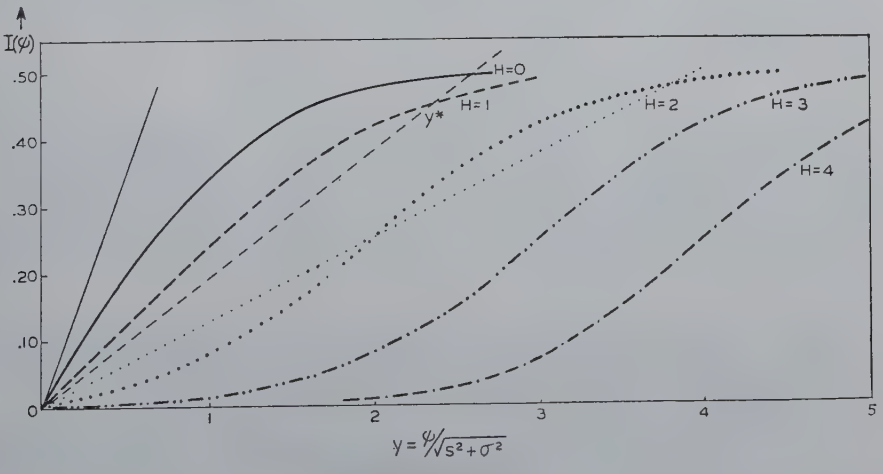

FIGURE 1

$$y = \psi/\sqrt{s^2 + \sigma^2}, \quad \boldsymbol{H} = h/\sqrt{s^2 + \sigma^2}, \tag{12}$$

and set $x_0 = 0$, we obtain from equations (10) and (11)

$$\begin{aligned} I(\sqrt{s^2 + \sigma^2}\, y_e) &= \tfrac{1}{2}G(y_e + \boldsymbol{H}) \\ &\quad + \tfrac{1}{2}G(y_e - \boldsymbol{H}) = \frac{a\sqrt{s^2 + \sigma^2}}{2AN_0}\, y_e. \end{aligned} \tag{13}$$

The expression $\frac{1}{2}G(y + \boldsymbol{H}) + \frac{1}{2}G(y - \boldsymbol{H})$ may be considered to be a function of y with $\boldsymbol{H}$ as a parameter. Figure 1 shows various members from this family of curves. The right-hand side of equation (13) is a straight line. Depending upon its slope and the value of $\boldsymbol{H}$, there may be one of the following ways in which the straight line intersects any particular one of the family of curves:

a) one intersection at the origin if $\boldsymbol{H}$ is small but

$$\sqrt{s^2 + \sigma^2}\, a/2AN_0 \text{ is large;}$$

b) three intersections if either $\boldsymbol{H}$ or $\sqrt{s^2 + \sigma^2}\, a/2AN_0$ is sufficiently small;

c) five intersections if $\boldsymbol{H}$ is sufficiently large and if

$$\sqrt{s^2 + \sigma^2}\, a/2AN_0$$

lies within certain limits.

The slope of I at $y = 0$ is equal to $g(\boldsymbol{H})$. The maximum value m of the slope of the line, tangent to the curve I and passing through the origin, can be expressed, approximately, for $\boldsymbol{H} \geq 1$ by the empirical expression

$$m = 1/(3.8 + 2\mathrm{H}) + 0.515e^{-2\mathrm{H}}. \tag{14}$$

It is clear that if m is greater than the slope of I at $y = 0$, which occurs if, and only if, $\boldsymbol{H} > 1$, it is possible for a straight line through the origin to intersect the curve I twice to the right of the origin. With the approximation (14) we may summarize the results as follows:

a) if $\boldsymbol{H} \leq 1$ and $\sqrt{s^2 + \sigma^2}\, a/2AN_0 \geq g(\boldsymbol{H})$, or if $\boldsymbol{H} > 1$ and $\sqrt{s^2 + \sigma^2}\, a/2AN_0 \geq m > g(\boldsymbol{H})$, there is but one point of equilibrium. It is stable at the origin (cf. solid curves in Figure 1);

b) if $\sqrt{s^2 + \sigma^2}\, a/2AN_0 < g(\boldsymbol{H})$, there is an unstable equilibrium at the origin and two stable equilibria at $\pm\, y_e$, which can be

calculated from equation (13) (cf. broken curves in Figure 1);

c) if $\boldsymbol{H} > 1$ and if $m > \sqrt{s^2 + \sigma^2}\, a/2AN_0 > g(\boldsymbol{H})$, there is a stable equilibrium at the origin, two unstable equilibria at $\pm y_e^*$ and two stable equilibria at $\pm y_e^{**}$, $y_e^{**} > y_e^*$. Both y_e^* and y_e^{**} satisfy equation (13) (cf. dotted curves in Figure 1).

If $x_0 \gtrless 0$, it is clear that the equilibrium at the origin will be displaced to an extent depending on the value of x_0, and the other possible roots of equation (13) will no longer be symmetrically placed about the origin.

If the distribution of Δ is approximated by an absolute value exponential function, results similar to those above are readily obtained.

Variable Threshold, No Bias. We consider next the case in which the threshold h varies from individual to individual and is distributed for the whole population according to some distribution function $N_0 q(h)$. For this case we neglect any individual bias and thus set $\phi = 0$ for each member of the population. The number X of individuals exhibiting R_1 is now

$$X = N_0 \int_0^\infty q(h) \int_{-\psi+h}^\infty p_1(\Delta)\, d\Delta\,. \tag{15}$$

We shall consider the special case in which $p_1(\Delta)$ and $q(h)$ are given by the following expressions:

$$p_1(\Delta) = \frac{k}{2} e^{-k|\Delta|}, \tag{16}$$

$$q(h) = \frac{ab}{b-a} (e^{-ah} - e^{-bh})\,. \tag{17}$$

We then find the following expression for $I(\psi)$ when $\psi > 0$:

$$I(\psi) = \tfrac{1}{4} \int_0^\infty q(h) \left\{ \int_{-\psi+h}^\infty k e^{-k|\Delta|} d\Delta - \int_{-\infty}^{-\psi-h} k e^{-k|\Delta|} d\Delta \right\} dh\,, \tag{18}$$

or

$$\begin{aligned} 2I(\psi) = 1 &+ \frac{b}{(k^2 - a^2)(b-a)} (a^2 e^{-k\psi} - k^2 e^{-a\psi}) \\ &+ \frac{a}{(k^2 - b^2)(b-a)} (k^2 e^{-b\psi} - b^2 e^{-k\psi})\,. \end{aligned} \tag{19}$$

Now if one of the parameters a or b in $q(h)$ is much larger than the other, e.g. $a >> b$, $q(h)$ is a simple exponential function be^{-bh}, and

$$I_1(\pm\psi) = \pm\tfrac{1}{2}|\{k^2(1-e^{-b|\psi|}) - b^2(1-e^{-k|\psi|})\}/(k^2-b^2)|, \quad (a >> b). \tag{20}$$

Thus $I_1'(\psi) > 0$, $I_1'(0) = kb/2(k+b)$, while $I_1''(\psi) < 0$ for $\psi > 0$. Thus

a) if $as/2AN_0 > kb/2(k+b)$, the origin is the only equilibrium point and it is stable;

b) if $as/2AN_0 < kb/2(k+b)$, the origin is an unstable equilibrium point and there are two stable equilibria at $\pm\psi_e$, where $I_1(\psi_e) = as\psi_e/2AN_0$.

If $a = b$, so that $q(h) = b^2he^{-bh}$, we find

$$2I_2(\psi) = 1 + \frac{(3b^2 - k^2 + b^3\psi - k^2b\psi)}{(k^2-b^2)^2}k^2e^{-b\psi} - \frac{b^2(k^2+b^2)}{(k^2-b^2)^2}e^{-k\psi}. \tag{21}$$

Furthermore, if $b = k$, it is readily shown that $I_2'(0) = b/8$ and $I_2''(1/k) = 0$; hence there is a point of inflection at $1/k$. More generally, we find that unless a or b becomes infinite there is an inflection point. When $b > k$ the inflection is more pronounced and the results of plotting $I(\psi)$ versus $b\psi$ for various values of k are similar to the results in the first case with constant h, but with variable bias.

If $I'' > 0$ in the neighborhood of the origin, then there will be an inflection point and it will be possible to have five equilibrium points. The differentiation can be carried out before integrating equation (18). The function I will be found to be concave upward near the origin if

$$q(0) < k\int_0^\infty q(h)e^{-kh}dh. \tag{22}$$

On the other hand, if $p(\Delta)$ is the normal distribution $g(h/s)$ and if $q(h)$ is arbitrary, then the function I will be concave upward in the neighborhood of the origin if

$$\int_0^\infty g(h/s)\,(h^2/s^2-1)\,q(h)dh > 0. \tag{23}$$

It may be noted that one would expect the variability in the threshold (h) or in the individual bias (ϕ) to be greater than the variability in the fluctuations. Thus, in this last case, the occurrence of a pronounced inflection would seem to be rather likely.

The author wishes to express his appreciation to Drs. N. Rashevsky and H. G. Landau for reading and discussing the manuscript.

LITERATURE

Landahl, H. D. 1938. "A Contribution to the Mathematical Biophysics of Psychophysical Discrimination." *Psychometrika*, **3**, 107–125.

Landau, H. G. 1950. "Note on the Effect of Imitation in Social Behavior." *Bull. Math. Biophysics*, **12**, 221–236.

Rashevsky, N. 1949. "Mathematical Biology of Social Behavior: III." *Bull. Math. Biophysics*, **11**, 255–271.

BULLETIN OF
MATHEMATICAL BIOPHYSICS
VOLUME 12, 1950

PSYCHOPHYSICAL DISCRIMINATION WITH MORE THAN TWO STIMULI

N. RASHEVSKY
COMMITTEE ON MATHEMATICAL BIOLOGY
THE UNIVERSITY OF CHICAGO

A generalization is suggested of H. D. Landahl's theory of psychophysical discrimination by considering instead of two cross-inhibiting parallel chains of neural pathways, n such chains. The probabilities of the n possible different reactions are in such a case expressed by $(n-1)$ple integrals. For $n=3$ the method is illustrated by evaluating the probability of a reaction to the weakest of the three stimuli.

As a generalization of H. D. Landahl's (1938; Rashevsky, 1948, chap. xxxiv; hereinafter referred to as MB) circuit of two parallel cross-inhibitory pathways, consider the case of three such pathways. Let the constants of the cross-inhibitory pathways be such that when the excitations ε_1, ε_2 and ε_3 of the three proximal connections are equal, the inhibitory effects just compensate the excitatory. This requires that the amount of j produced by the cross-inhibitory pathway, l, m, leading from the lth pathway to the mth $(l, m = 1, 2, 3)$ be equal to $\varepsilon_l/2$.

Consider the case in which

$$\frac{\varepsilon_1 + \varepsilon_2}{2} - \varepsilon_3 = \Delta > 0. \tag{1}$$

In the absence of fluctuations the reaction R_3 in the third pathway will not occur when (1) holds. Due to fluctuations at the connections it may, however, happen that the excitation at the third connection will exceed the half-sum of the excitations at the first and second connections, and a reaction R_3 will occur accidentally, as a "wrong" reaction. If ϕ_1, ϕ_2 and ϕ_3 denote the accidental additional amounts of excitation at the three corresponding connections, due to fluctuations, then in order that R_3 will occur, we must have

$$\frac{\varepsilon_1 + \phi_1 + \varepsilon_2 + \phi_2}{2} - (\varepsilon_3 + \phi_3) < 0, \tag{2}$$

or

$$\phi_3 > \Delta + \frac{\phi_1 + \phi_2}{2}. \tag{3}$$

Let the probability that any of the ϕ_i's will have a value between x and $x + dx$ be $p(x)dx$. That is, the probability of ϕ_1 having a value between ϕ_1 and $\phi_1 + d\phi_{11}$ is $p(\phi_1)d\phi_1$, while the probability of ϕ_2 having a value between ϕ_2 and $\phi_2 + d\phi_2$ is $p(\phi_2)d\phi_2$, etc. Then the probability of inequality (3), when the values ϕ_1 and ϕ_2 are known, is

$$\int_{\Delta+(\phi_1+\phi_2)/2}^{\infty} p(x)dx = F(\Delta, \phi_1, \phi_2). \tag{4}$$

The probability that ϕ_1 and ϕ_2 have prescribed specific values is

$$p(\phi_1)p(\phi_2)d\phi_1 d\phi_2. \tag{5}$$

Therefore, the probability that inequality (3) holds for specified values ϕ_1 and ϕ_2 is given by

$$p(\phi_1)p(\phi_2)F(\Delta, \phi_1, \phi_2)d\phi_1 d\phi_2. \tag{6}$$

Hence the probability that (3) will hold for any values of ϕ_1 and ϕ_2, that is, the probability P_w of R_3, is given by

$$P_w(\Delta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(\phi_1)p(\phi_2)F(\Delta, \phi_1, \phi_2)d\phi_1 d\phi_2. \tag{7}$$

A similar procedure may be applied to the case of n stimuli, and we obtain an expression for P_w in terms of an $(n-1)$ple integral.

In order to compute $P_w(\Delta)$ explicitly, we must specify the function $p(x)$. For the same reasons as before we shall use the following expression (Landahl, 1938)

$$p(x) = \frac{k}{2}e^{-k|x|}. \tag{8}$$

The function $p(x)$ has two different expressions, according to whether $x > 0$ or $x < 0$. We have

$$\text{For } x > 0 \qquad p(x) = \overset{+}{p}(x) = \frac{k}{2}e^{-kx}; \tag{9}$$

$$\text{For } x < 0 \qquad p(x) = \overset{-}{p}(x) = \frac{k}{2}e^{kx}. \tag{10}$$

In evaluating $F(\Delta, \phi_1, \phi_2)$ we obtain different expressions, depending on whether $\frac{1}{2}(\phi_1 + \phi_2) > -\Delta$ or $\frac{1}{2}(\phi_1 + \phi_2) < -\Delta$. In the

first instance we have

$$\overset{+}{F}(\Delta, \phi_1, \phi_2) = \frac{k}{2} \int_{\Delta+(\phi_1+\phi_2)/2}^{\infty} e^{-kx} dx$$
$$= \tfrac{1}{2}\, e^{-k\left(\Delta + \frac{\phi_1+\phi_2}{2}\right)}. \tag{11}$$

In the second case we must divide the range of integration into a negative and positive part, and find

$$\overset{-}{F}(\Delta, \phi_1, \phi_2) = \frac{k}{2} \int_{\Delta+(\phi_1+\phi_2)/2}^{\infty} e^{-k|x|} dx$$
$$= \frac{k}{2} \int_{\Delta+(\phi_1+\phi_2)/2}^{0} e^{kx} dx + \frac{k}{2} \int_{0}^{\infty} e^{-kx} dx \tag{12}$$
$$= 1 - \tfrac{1}{2}\, e^{k\left(\Delta + \frac{\phi_1+\phi_2}{2}\right)}.$$

In evaluating the integral (7) we must break up the range of integration according to the values $\overset{+}{p}(x)$, $\overset{-}{p}(x)$, $\overset{+}{F}$, and $\overset{-}{F}$ of the functions $p(x)$ and $F(\Delta, \phi_1, \phi_2)$.

First we break it according to the values of $p(x)$. We have from (7)

$$P_w(\Delta) = \int_{-\infty}^{+\infty} p(\phi_2)\, d\phi_2 \int_{-\infty}^{0} \overset{-}{p}(\phi_1) F(\Delta, \phi_1, \phi_2)\, d\phi_1$$
$$+ \int_{-\infty}^{+\infty} p(\phi_2)\, d\phi_2 \int_{0}^{\infty} \overset{+}{p}(\phi_1) F(\Delta, \phi_1, \phi_2)\, d\phi_1 . \tag{13}$$

In evaluating the integral

$$\int_{-\infty}^{0} \overset{-}{p}(\phi_1) F(\Delta, \phi_1, \phi_2)\, d\phi_1 \tag{14}$$

we must find the relation between ϕ_1 and ϕ_2 within the range of negative values of ϕ_1. We have seen [equations (10) and (12)] that

$$F = \overset{+}{F} \quad \text{for} \quad \phi_1 + \phi_2 > -2\Delta;$$
$$F = \overset{-}{F} \quad \text{for} \quad \phi_1 + \phi_2 < -2\Delta. \tag{15}$$

In the ϕ_1, ϕ_2 plane draw the straight line $\phi_1 + \phi_2 = -2\Delta$.

(Figure 1.) Above the line $F = F$; below the line $F = \bar{F}$.

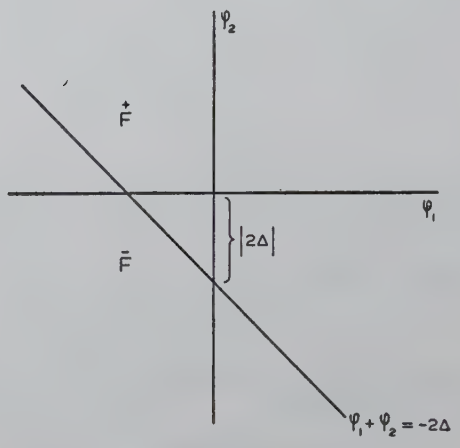

FIGURE 1

If $\phi_2 > -2\Delta$, then $F = \bar{F}$ within the range of ϕ_1 from $-\infty$ to $-(2\Delta + \phi_2)$; while for ϕ_1 between $-(2\Delta + \phi_2)$ and 0, $F = \overset{+}{F}$. If $\phi_2 < -2\Delta$, then $F = \bar{F}$ for all negative values of ϕ_1. Hence the integral (14) can be written as follows.

For $\phi_2 > -2\Delta$:

$$\int_{-\infty}^{0} \bar{p}(\phi_1) F d\phi_1 = \int_{-\infty}^{-(2\Delta+\phi_2)} \bar{p}(\phi_1) \bar{F} d\phi_1 + \int_{-(2\Delta+\phi_2)}^{0} \bar{p}(\phi_1) \overset{+}{F} d\phi_1; \quad (16)$$

For $\phi_2 < -2\Delta$:

$$\int_{-\infty}^{0} \bar{p}(\phi_1) F d\phi_1 = \int_{-\infty}^{0} \bar{p}(\phi_1) \bar{F} d\phi_1. \quad (17)$$

By a similar argument we find for the last integral of equation (13):

For $\phi_2 > -2\Delta$:

$$\int_{0}^{\infty} \overset{+}{p}(\phi_1) F d\phi_1 = \int_{0}^{\infty} \overset{+}{p}(\phi_1) \overset{+}{F} d\phi_1. \quad (18)$$

For $\phi_2 < -2\Delta$:

$$\int_{0}^{\infty} \overset{+}{p}(\phi_1) F d\phi_1 = \int_{0}^{-(2\Delta+\phi_2)} \overset{+}{p}(\phi_1) \bar{F} d\phi_1 + \int_{-(2\Delta+\phi_2)}^{\infty} \overset{+}{p}(\phi_1) \overset{+}{F} d\phi_1. \quad (19)$$

Because of (9) and (10) the range of integration with respect to ϕ_2 in the first term of the right-hand side of (13) must be broken into $(-\infty, 0)$ and $(0, +\infty)$. Because of (16) and (17), the range $(-\infty, 0)$ must again be broken into two: $(-\infty, -2\Delta)$ and $(-2\Delta, 0)$. Keeping this in mind, we find for the first term of the right-hand side of (13)

$$\int_{-\infty}^{-2\Delta} \bar{p}(\phi_2)\,d\phi_2 \int_{-\infty}^{0} \bar{p}(\phi_1)\bar{F}d\phi_1 + \int_{-2\Delta}^{0} \bar{p}(\phi_2)\,d\phi_2 \left\{ \int_{-\infty}^{-(2\Delta+\phi_2)} \bar{p}(\phi_1)\bar{F}d\phi_1 + \int_{-(2\Delta+\phi_2)}^{0} \bar{p}(\phi_1)\overset{+}{F}d\phi_1 \right\} + \int_{0}^{\infty} \overset{+}{p}(\phi_2)\,d\phi_2 \left\{ \int_{-\infty}^{-(2\Delta+\phi_2)} \bar{p}(\phi_1)\bar{F}d\phi_1 + \int_{-(2\Delta+\phi_2)}^{0} \bar{p}(\phi_1)\overset{+}{F}d\phi_1 \right\}. \tag{20}$$

In a similar way we break up the second term of the right-hand side of (13). Using (9), (10), (16), (17), (18) and (19) and evaluating all the integrals in (20) and in the corresponding expression for the second term of the right side of (13), we finally obtain

$$P_w(\Delta) = \frac{8}{9}e^{-k\Delta} - \left(\frac{k\Delta}{6} + \frac{7}{18}\right)e^{-2k\Delta}. \tag{21}$$

For $\Delta = 0$, $P_w = \frac{1}{2}$. This at first seems paradoxical. When $\varepsilon_1 = \varepsilon_2 = \varepsilon_3$ we have $\Delta = 0$, and we would expect that because of complete symmetry the probability of any of the three reactions would be $\frac{1}{3}$. We must remember, however, that whereas in the case of two stimuli only one of the reactions may occur, in the present case this is not necessarily so. Therefore $P_w(0)$ means the probability for $\varepsilon_1 = \varepsilon_2 = \varepsilon_3$ of R_3 to occur either alone or in combination with either R_1 or R_2, but not with both.

Let us compare equation (21) with the one which we obtain if we assume that one of the stimuli — say, ε_2 — is entirely cut out. By this we do not mean that $\varepsilon_2 = 0$, for that case is covered by the preceding argument. We actually consider one of the pathways as not functioning, so that now we have

$$\frac{\varepsilon_1}{2} - \varepsilon_3 = \Delta > 0. \tag{22}$$

We may now compute $P_w'(\Delta)$ of the reaction R_3.

The condition for R_3 to occur is now

$$\phi_3 > \frac{\phi_1}{2} + \Delta, \tag{23}$$

the probability of which is

$$\int_{\Delta+\phi_1/2}^{\infty} p(x)\,dx = F(\Delta, \phi_1). \tag{24}$$

The total probability $P_w(\Delta)$ is given by

$$P_w'(\Delta) = \int_{-\infty}^{+\infty} F p d\,\phi_1 \tag{25}$$

and is readily computed in the same way as before. We find

$$P_w'(\Delta) = \frac{2}{3} e^{-k\Delta} - \frac{1}{6} e^{-2k\Delta}. \tag{26}$$

It is readily seen that for large values of Δ, when the terms in $e^{-2k\Delta}$ become small compared with those in $e^{-k\Delta}$, $P_w'(\Delta) < P_w(\Delta)$.

It is also readily seen that for small values of Δ $P_w'(\Delta) < P_w(\Delta)$. We have $P_w'(0) = P_w(0) = \frac{1}{2}$. Furthermore for $\Delta = 0$ we have $dP_w'/d\Delta = -k/3$; $dP_w/d\Delta = -5k/18$.

The author is indebted to Mr. George Karreman for checking the manuscript.

LITERATURE

Landahl, H. D. 1938. "A Contribution to the Mathematical Biophysics of Psychophysical Discrimination." *Psychometrika*, 3, 107–125.

Rashevsky, N. 1948. *Mathematical Biophysics.* Rev. Ed. Chicago: University of Chicago Press.

BULLETIN OF
MATHEMATICAL BIOPHYSICS
VOLUME 12, 1950

NOTE ON THE EFFECT OF IMITATION IN SOCIAL BEHAVIOR

H. G. LANDAU
COMMITTEE ON MATHEMATICAL BIOLOGY
THE UNIVERSITY OF CHICAGO

The discussion given by N. Rashevsky (1949) on the effect of imitation in the mathematical biology of social behavior is generalized by assuming the distributions involved to be normal rather than Laplace distributions, and also by showing how most of the results can be derived without assuming any specific form for the distributions. In particular, it is demonstrated that it is possible, in a sufficiently large population, to have a stable behavior pattern which is quite independent of the desires of the population or of their inherent pattern of response.

Introduction.

N. Rashevsky (1949, in the following referred to as R) considered the effect of imitation in a population where each individual could respond to a given stimulus with one of two possible reactions, R_1 or R_2. Using the notation of R, each individual is characterized by ϕ, the net excitation to response R_1, where ϕ is distributed in the population so that the number of individuals whose ϕ is between ϕ and $\phi + d\phi$ is given by

$$N(\phi)\,d\phi = N_0 n(\phi)\,d\phi\,, \tag{1}$$

where N_0 is the total number of individuals in the population, and

$$\int_{-\infty}^{\infty} n(\phi)\,d\phi = 1\,. \tag{2}$$

In the absence of imitation, an individual responds with reaction R_1 if $\phi + \phi' > 0$ (and R_2 if $\phi + \phi' < 0$), where ϕ' is a random excitation. This idea was introduced by H. D. Landahl (1938) in his treatment of psychophysical discrimination. The probability density function of ϕ' is $p(\phi')$, so that the probability of reaction R_1 for an individual ϕ is

$$P_1(\phi) = \int_{-\phi}^{\infty} p(\phi')\,d\phi'. \tag{3}$$

It is assumed that the effect of imitation is to add an additional excitation, the stimulus toward R_1 being proportional to X, the number of individuals exhibiting R_1, and the corresponding stimulus toward R_2 being proportional to $N_0 - X$. These excitations due to imitation are assumed to follow the usual laws for neural excitation (Rashevsky, 1948), so that the net excitation to R_1 due to imitation, ψ, is given by the differential equation

$$\frac{d\psi}{dt} = A(2X - N_0) - a\psi, \tag{4}$$

the positive parameters, A and a, being assumed to have the same value for all members of the population. The quantity X is given by

$$X(\psi) = \int_{-\infty}^{\infty} P_1(\phi + \psi)\, N(\phi)\, d\phi\,. \tag{5}$$

In R it was assumed that $n(\phi)$ and $p(\phi')$ were Laplace distributions, i.e.

$$n(\phi) = \frac{\sigma}{2} e^{-\sigma|\phi|}$$

and

$$p(\phi') = \frac{k}{2} e^{-k|\phi'|},$$

and the nature and stability of the solutions of (4) were studied. Here similar results are derived assuming $n(\phi)$ and $p(\phi')$ to be normal distributions. This is not only a more plausible form but the calculations are simpler. It is also shown that most of the conclusions may be established without assuming a specific form for the distributions.

Normal Distribution: Unbiased Population.

We use the following notation for the normal distribution and its integral

$$\begin{aligned} g(x) &= (2\pi)^{-\frac{1}{2}} e^{-\frac{x^2}{2}}\,, \\ G(x) &= \int_0^x g(x')\, dx'\,. \end{aligned} \tag{6}$$

If we now take for $n(\phi)$ and $p(\phi')$

$$n(\phi) = \sigma^{-1} g\left(\frac{\phi}{\sigma} \right),$$
$$p(\phi') = s^{-1} g\left(\frac{\phi'}{s} \right), \tag{7}$$

then we obtain from (3) and (5),

$$X(\psi) = \frac{N_0}{\sigma} \int_{-\infty}^{\infty} \left[\tfrac{1}{2} + G\left(\frac{\phi + \psi}{s} \right) \right] g\left(\frac{\phi}{\sigma} \right) d\phi. \tag{8}$$

Writing (4) as

$$\frac{d\psi}{dt} = 2AN_0 I(\psi) - a\psi = F(\psi), \tag{9}$$

so that

$$I(\psi) = \frac{X(\psi)}{N_0} - \frac{1}{2}, \tag{10}$$

we have from (8)

$$I(\psi) = \sigma^{-1} \int_{-\infty}^{\infty} G\left(\frac{\phi + \psi}{s} \right) g\left(\frac{\phi}{\sigma} \right) d\phi. \tag{11}$$

This definite integral is easily evaluated. Putting $x = \phi/\sigma$, $\alpha = \sigma/s$, $\beta = \psi/s$ in (10), we obtain

$$I = \int_{-\infty}^{\infty} G(\alpha x + \beta)\, g(x)\, dx \tag{12}$$

and

$$\frac{\partial I}{\partial \beta} = \int_{-\infty}^{\infty} g(\alpha x + \beta)\, g(x)\, dx$$
$$= g\left(\frac{\beta}{\sqrt{1+\alpha^2}} \right) \int_{-\infty}^{\infty} g\left(\sqrt{1+\alpha^2}\, x + \frac{\alpha\beta}{\sqrt{1+\alpha^2}} \right) dx$$
$$= (1+\alpha^2)^{-\frac{1}{2}} g\left(\frac{\beta}{\sqrt{1+\alpha^2}} \right) \int_{-\infty}^{\infty} g(\xi)\, d\xi$$
$$= (1+\alpha^2)^{-\frac{1}{2}} g\left(\frac{\beta}{\sqrt{1+\alpha^2}} \right).$$

Hence

$$I = G\left(\frac{\beta}{\sqrt{1+\alpha^2}}\right) + f(\alpha).$$

Since $G(x)$ is an odd function and $g(x)$ is an even function, $I = 0$ for $\beta = 0$, and hence $f(\alpha) = 0$. This gives

$$I = G\left(\frac{\beta}{\sqrt{1+\alpha^2}}\right), \tag{13}$$

and therefore

$$I(\psi) = G\left(\frac{\psi}{\sqrt{s^2+\sigma^2}}\right). \tag{14}$$

The differential equation (9) for ψ now becomes

$$\frac{d\psi}{dt} = 2AN_0\, G\left(\frac{\psi}{\sqrt{s^2+\sigma^2}}\right) - a\psi = F(\psi). \tag{15}$$

The behavior of the possible solutions of (15) can now be determined. We first note that $F(0) = 0$, so that $\psi = 0$ is a possible solution. This will be stable, in the sense that if ψ is changed by a small amount to a value different from zero it will return to zero, whenever $dF/d\psi < 0$ at $\psi = 0$; that is,

$$\frac{2AN_0}{\sqrt{2\pi(s^2+\sigma^2)}} - a < 0$$

or

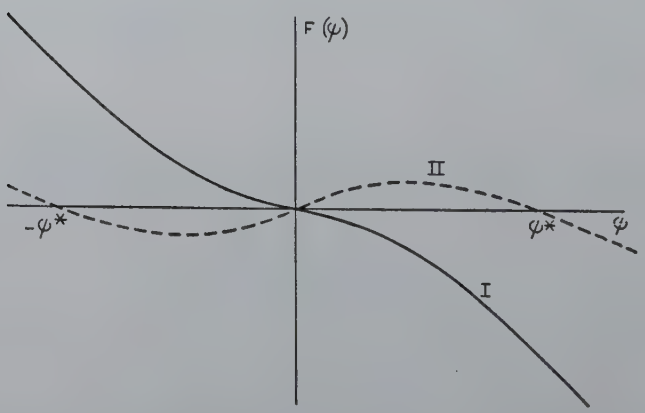

FIGURE 1. The alternatives for $F(\psi)$ for normal distributions, unbiased population.

$$\gamma = \frac{a\sqrt{s^2+\sigma^2}}{2AN_0} > (2\pi)^{-\frac{1}{2}} = .3989. \tag{16}$$

In this case

$$\frac{dF}{d\psi} = \frac{2AN_0}{\sqrt{s^2+\sigma^2}} g\left(\frac{\psi}{\sqrt{s^2+\sigma^2}}\right) - a = a\left[\frac{1}{\gamma} g\left(\frac{\psi}{\sqrt{s^2+\sigma^2}}\right) - 1\right] \tag{17}$$

will be negative for all values of ψ, so that $d\psi/dt > 0$ for $\psi < 0$ and $d\psi/dt < 0$ for $\psi > 0$. This means that no matter what value ψ may have initially, it tends to return to the value of $\psi = 0$. This is the situation shown in Curve I of Figure 1. Since $\psi = 0$ gives $X = N_0/2$, this means that a population characterized by values of the parameter in this range tends to be equally divided between reaction R_1 and R_2.

If the inequality (16) is reversed to

$$\gamma < (2\pi)^{-\frac{1}{2}}, \tag{18}$$

then $\psi = 0$ will not be a stable solution. Suppose ψ is increased to a small positive value, we have $F(\psi) > 0$ and from equation (15) it can be seen that ψ will continue to increase up to the value for which $F(\psi) = 0$. That a positive root ψ^* of $F(\psi) = 0$ exists can be seen from the fact that

$$G\left(\frac{\psi}{\sqrt{s^2+\sigma^2}}\right) < \tfrac{1}{2}$$

while $a\psi$ increases indefinitely. This is the situation shown in Curve II of Figure 1. Since

$$g\left(\frac{\psi}{\sqrt{s^2+\sigma^2}}\right)$$

is monotone decreasing for $\psi > 0$, it follows from (17) that there is only one positive root, ψ^*. By symmetry, there is a negative root of equal absolute value $-\psi^*$. From (17) and (18), we see that $dF(0)/d\psi > 0$ and decreases monotonically for $\psi > 0$, hence $dF(\psi^*)/d\psi < 0$, so that ψ^* is a stable value of ψ, and similarly $-\psi^*$ is also stable.

Hence we have the alternatives: either (16) holds, $\psi = 0$ is stable, and the population is equally divided between R_1 and R_2 and always returns to this configuration after any fluctuation; or (18) holds, $\psi = \psi^*$ and $\psi = -\psi^*$ are stable, and any positive fluctuation causes the population to move toward and then remain at $\psi = \psi^*$

with an excess of R_1 over R_2 (or a negative fluctuation sends it to the stable value $\psi = -\psi^*$ with an excess of R_2 over R_1). The equality

$$\gamma = (2\pi)^{-\frac{1}{2}} \tag{19}$$

can also be seen to give only $\psi = 0$ as a stable value.

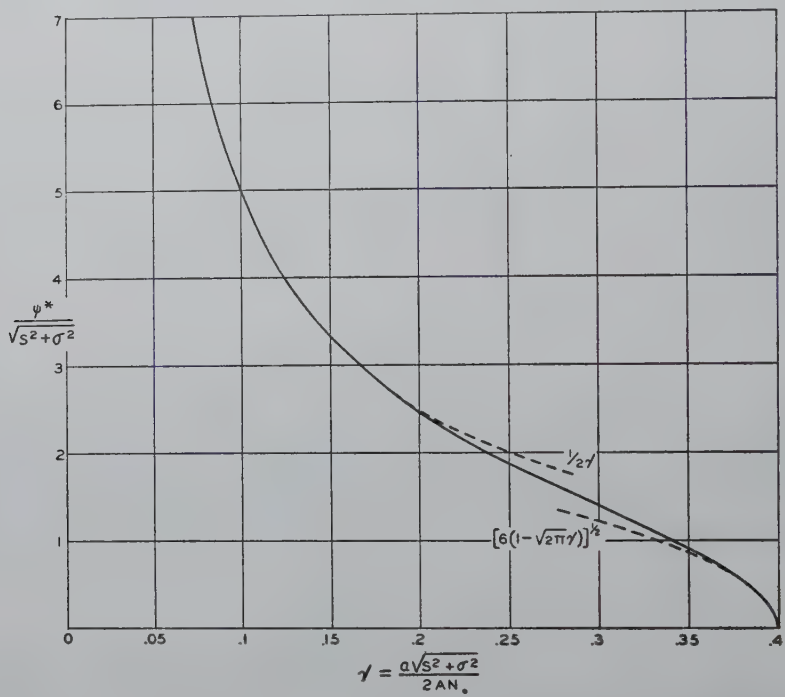

FIGURE 2. The roots of $F(\psi) = 0$ for normal distributions, unbiased population.

Numerical values of ψ^* can be calculated as follows. If we put $\xi = \psi/\sqrt{s^2 + \sigma^2}$, $F(\psi) = 0$ becomes $G(\xi)/\xi = \gamma$. Suppose $y = G(x)/x$ and let $x = H(y)$ be the inverse function for positive values of x; then $\xi = \psi^*/\sqrt{s^2 + \sigma^2} = H(\gamma)$. This relation is graphed in Figure 2. For small γ, we have approximately

$$\psi^* \cong \frac{\sqrt{s^2 + \sigma^2}}{2\gamma} = \frac{AN_0}{a}. \tag{20}$$

Since γ is small for large N_0, equation (20) would hold for a large population. At the other end of the range we can also obtain a simple approximate formula for ψ^*. Let

$$\gamma = (2\pi)^{-\frac{1}{2}}(1 - \varepsilon), \tag{21}$$

where ε is a small positive quantity; then

$$\psi^* = [6\varepsilon(s^2 + \sigma^2)]^{\frac{1}{2}}. \tag{22}$$

The upper and lower dashed lines in Figure 2 show the accuracy of these approximations.

It should be noted that the value of ψ^* in (20) does not depend on the distribution of ϕ (i.e. the parameters σ and s), but only on A, a and N_0. These parameters are determined by the effect of the tendency to imitate and the population size.

Normal Distribution: Biased Population.

Suppose the population is biased in its distribution of ϕ, the characteristic net excitation to response R_1. We assume that instead of (7) the distribution is given by

$$n(\phi) = \sigma^{-1} g\left(\frac{\phi - \phi_0}{\sigma}\right), \tag{23}$$

where $\phi_0 > 0$ so that the population is biased in favor of R_1. Then we have

$$X(\psi) = \frac{N_0}{\sigma} \int_{-\infty}^{\infty} \left[\tfrac{1}{2} + G\left(\frac{\phi + \psi}{s}\right)\right] g\left(\frac{\phi - \phi_0}{\sigma}\right) d\phi, \tag{24}$$

and

$$\begin{aligned} I(\psi) &= \sigma^{-1} \int_{-\infty}^{\infty} G\left(\frac{\phi + \psi}{s}\right) g\left(\frac{\phi - \phi_0}{\sigma}\right) d\phi \\ &= \sigma^{-1} \int_{-\infty}^{\infty} G\left(\frac{\phi_1 + \psi + \phi_0}{s}\right) g\left(\frac{\phi_1}{\sigma}\right) d\phi_1, \end{aligned} \tag{25}$$

so that from equation (14) it follows that

$$I(\psi) = G\left(\frac{\psi + \phi_0}{\sqrt{s^2 + \sigma^2}}\right). \tag{26}$$

The differential equation for ψ now becomes

$$\frac{d\psi}{dt} = 2AN_0 G\left(\frac{\psi + \phi_0}{\sqrt{s^2 + \sigma^2}}\right) - a\psi = F(\psi). \tag{27}$$

In this case $F(0) > 0$, but it is easy to see that there will always be a positive root $F(\psi^*) = 0$ for $\psi^* > 0$ because

$$\left| G\left(\frac{\psi + \phi_0}{\sqrt{s^2 + \sigma^2}}\right) \right| < \tfrac{1}{2},$$

while $a\psi$ increases indefinitely as ψ increases. There is only one positive root and it is stable since

$$\frac{dF}{d\psi} = a\left[\frac{1}{\gamma} g\left(\frac{\psi + \phi_0}{\sqrt{s^2 + \sigma^2}}\right) - 1\right] \tag{28}$$

decreases with increasing ψ for $\psi > 0$, and this together with $F(0) > 0$ means that there can be only one positive root $F(\psi^*) = 0$ and that $dF(\psi^*)/d\psi < 0$. This is the situation in Curve I of Figure 3.

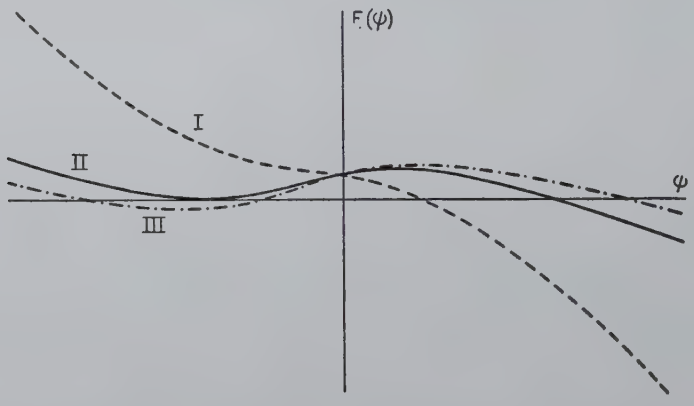

FIGURE 3. The three possibilities for $F(\psi)$ for normal distributions, biased population.

If inequality (16), $\gamma > (2\pi)^{-\frac{1}{2}}$, holds, then from (28) we see that $dF/d\psi < 0$ for all values of ψ, and the positive root is the only one. For $\gamma < (2\pi)^{-\frac{1}{2}}$ it is possible, depending on the value of ϕ_0, to have one or two negative roots. The condition under which these possibilities occur is derived below. That these are the only two possibilities can be seen from equation (28), which shows that $dF/d\psi = 0$ for at most one negative value of ψ, and must be negative for large negative values of ψ, since $dF/d\psi \to -a$ as $\psi \to -\infty$. Hence we have the two situations shown in II and III of Figure 3. In II, $F(\psi) = 0$, and $dF/d\psi = 0$ for the same value of ψ. Such a root is unstable. In III, there are two negative roots and the one with the smaller absolute value is unstable, but the other is stable.

The condition III can occur for sufficiently small γ for any value of ϕ_0. If we interpret the bias ϕ_0 as expressing the desire of the population for R_1, this means that in a sufficiently large population

it is possible to have a stable condition in which the majority behaves contrary to its own desire because of the effect of imitation. An effect of this kind may be part of the explanation of how the cooperation of the people of a country is obtained in an unwanted war.

We can derive the condition for $F(\psi)$ to have negative roots as follows. From the discussion above it can be seen that γ must be smaller than the value for which $F(\psi) = 0$ and $dF/d\psi = 0$ have the same root. If we put

$$\xi = \frac{\psi}{\sqrt{s^2 + \sigma^2}}, \quad \text{and} \quad \xi_0 = \frac{\phi_0}{\sqrt{s^2 + \sigma^2}}, \tag{29}$$

then $F(\psi) = 0$ and $dF/d\psi = 0$ become

$$G(\xi + \xi_0) - \gamma\xi = 0, \tag{30}$$

$$g(\xi + \xi_0) - \gamma = 0. \tag{31}$$

Hence we obtain

$$G(\xi + \xi_0) = \xi g(\xi + \xi_0). \tag{32}$$

The root ξ of (30) and (31) gives the point at which $G(\xi + \xi_0)/\xi$ has a negative maximum, and γ must be less than this maximum for negative roots of (30) to exist.

To utilize (32) we need the following expansion of $G(x)$ (Rosser, 1948), which is obtained by successive integrations by parts:

$$\begin{aligned} G(x) &= \int_0^x g(t)\,dt = xg(x) + \int_0^x t^2 g(t)\,dt \\ &= xg(x) + \frac{x^3}{3} g(x) + \tfrac{1}{3} \int_0^x t^4 g(t)\,dt \\ &= g(x)\left[x + \frac{x^3}{1\cdot 3} + \frac{x^5}{1\cdot 3\cdot 5} + \frac{x^7}{1\cdot 3\cdot 5\cdot 7} + \cdots \right]. \end{aligned} \tag{33}$$

Therefore from (32) it follows that

$$\xi = \xi + \xi_0 + \frac{(\xi + \xi_0)^3}{1\cdot 3} + \frac{(\xi + \xi_0)^5}{1\cdot 3\cdot 5} + \cdots \tag{34}$$

or

$$\xi_0 = -\left[\frac{(\xi + \xi_0)^3}{1\cdot 3} + \frac{(\xi + \xi_0)^5}{1\cdot 3\cdot 5} + \cdots \right]. \tag{35}$$

Solving (31) for $\xi + \xi_0$ gives

$$\xi + \xi_0 = -(-\log 2\pi\gamma^2)^{\frac{1}{2}}. \tag{36}$$

Hence from (35) we find

$$\xi_0 = \left[\frac{1}{1\cdot 3}(-\log 2\pi\gamma^2)^{3/2} + \frac{1}{1\cdot 3\cdot 5}(-\log 2\pi\gamma^2)^{5/2} + \cdots \right]. \tag{37}$$

This is the condition for the existence of a single unstable root. For two negative roots to exist γ must be smaller; that is, the following inequality must hold

$$\xi_0 > \left[\frac{1}{1\cdot 3}(-\log 2\pi\gamma^2)^{3/2} + \frac{1}{1\cdot 3\cdot 5}(-\log 2\pi\gamma^2)^{5/2} + \cdots \right]. \tag{38}$$

For $\phi_0 = 0$, inequality (38) of course reduces to the previous inequality (18).

For small values of ξ_0, γ must be near $(2\pi)^{-\frac{1}{2}}$ to satisfy (37). Then we can use the first term of the series, and (38) becomes

$$\gamma < g((3\xi_0)^{1/3}). \tag{39}$$

For large values of ξ_0, $\gamma \to 0$ and (38) becomes

$$\gamma < \frac{1}{2\xi_0} \quad \text{or} \quad \frac{a}{AN_0} < \frac{1}{\phi_0}. \tag{40}$$

Formulae for the roots in this case would be rather complicated but it is easy to determine them graphically. Letting

$$\eta = \xi + \xi_0 = \frac{\psi + \phi_0}{\sqrt{s^2 + \sigma^2}}, \tag{41}$$

we need merely plot $G(\eta)$ against η as abscissa and on the same graph draw the line with ordinate $\gamma(\eta - \xi_0)$. If η^* is the abscissa of the point (or points) of intersection, then the root ψ^* is given by

$$\psi^* = \sqrt{s^2 + \sigma^2}\,\eta^* - \phi_0.$$

General Distributions.

In this section we show that many of the conclusions in R and in the previous sections can be established independently of any assumed specific form for the distributions $n(\phi)$ and $p(\phi')$. Equations (9) and (10) are independent of the form of these distributions. From the definition of $X(\psi)$ as the number of individuals exhibiting response R_1, it follows that

$$0 \leq X(\psi) \leq N_0 \tag{42}$$

and hence

$$-\tfrac{1}{2} \leq I(\psi) \leq \tfrac{1}{2}. \tag{43}$$

From the continuity of $I(\psi)$ and (9) it follows that $F(\psi) = 0$ has one or more roots, ψ^*, and that these roots lie in the interval

$$-\frac{AN_0}{a} \leq \psi^* \leq \frac{AN_0}{a}. \tag{44}$$

Furthermore, either $dF/d\psi$ must be negative at some root or else $dF(\psi^*)/d\psi = 0$ with $F(\psi) < 0$ for $\psi > \psi^*$ and $F(\psi) > 0$ for $\psi < \psi^*$†, so that there is a stable root.

The bounds for this root in (44) are the limiting values in equation (20). We can also show that as N_0 becomes large, there always are two roots which asymptotically approach these limits. From (3) and (5) it can be seen that as $\psi \to \infty$, $X(\psi) \to N_0$ and, consequently, $I(\psi) \to \frac{1}{2}$. That is, for any $\varepsilon > 0$ there is a $\psi_\varepsilon > 0$ such that $(1 - \varepsilon)/2 < I(\psi) \leq \frac{1}{2}$ for $\psi > \psi_\varepsilon$. Hence

$$F(\psi) = 2AN_0 I(\psi) - a\psi > 0 \quad \text{for} \quad \psi_\varepsilon < \psi < \frac{AN_0(1-\varepsilon)}{a},$$

and

$$F(\psi) \leq 0 \quad \text{for} \quad \psi \geq \frac{AN_0}{a},$$

from which it follows that for sufficiently large N_0 there must be a stable positive root ψ^* satisfying

$$\frac{AN_0}{a}(1-\varepsilon) < \psi^* \leq \frac{AN_0}{a}. \tag{45}$$

By considering $\psi \to -\infty$, we find a similar negative root.

The significance of this result lies in the demonstration that imitation makes it possible, in a sufficiently large population, to have a stable behavior pattern which is entirely independent of the desires of the population or of its inherent pattern of response.

We now assume that $n(\phi)$ and $p(\phi')$ are symmetric or even functions; that is

$$n(-\phi) = n(\phi) \text{ and } p(-\phi') = p(\phi),$$

so that

†There might be a whole interval of roots, $F(\psi) = 0$ for $\psi_1^* \leq \psi \leq \psi_2^*$, with $dF(\psi_2^*)/d\psi < 0$ and $dF(\psi_1^*)/d\psi < 0$. Then every point in the interval is a stable root.

$$\int_0^\infty n(\phi)\,d\phi = \tfrac{1}{2} \text{ and } P_1(-\phi) = 1 - P_1(\phi). \tag{46}$$

Then from (5) and (10) we find

$$I(\psi) = -\tfrac{1}{2} + \int_{-\infty}^{\infty} P_1(\phi + \psi)\, n(\phi)\, d\phi\,. \tag{47}$$

Dividing the integral into two parts and changing the sign of the integration variable we obtain

$$\begin{aligned} I(\psi) &= -\tfrac{1}{2} + \int_0^\infty P_1(\phi + \psi)\, n(\phi)\, d\phi + \int_{-\infty}^{0} P_1(\phi + \psi)\, n(\phi)\, d\phi \\ &= -\tfrac{1}{2} + \int_0^\infty P_1(\phi + \psi)\, n(\phi)\, d\phi + \int_0^\infty P_1(-\phi + \psi)\, n(-\phi)\, d\phi \\ &= -\tfrac{1}{2} + \int_0^\infty [P_1(\phi + \psi) + 1 - P_1(\phi - \psi)]\, n(\phi)\, d\phi\,. \end{aligned}$$

From the above we finally obtain

$$I(\psi) = \int_0^\infty [P_1(\phi + \psi) - P_1(\phi - \psi)]\, n(\phi)\, d\phi\,. \tag{48}$$

This shows that in the present case

$$I(\psi) = -I(-\psi) \tag{49}$$

and $I(0) = 0$. Hence there is a root $F(0) = 0$.

Now

$$I'(\psi) = \frac{dI}{d\psi} = \int_0^\infty [p(\phi + \psi) + p(\phi - \psi)]\, n(\phi)\, d\phi, \tag{50}$$

and

$$I'(0) = 2 \int_0^\infty p(\phi)\, n(\phi)\, d\phi > 0\,, \tag{51}$$

because $p(\phi) \geq 0$ and $n(\phi) \geq 0$ and neither can be identically zero. The root at $\psi^* = 0$ will be stable or unstable according to whether

$$F'(0) = 2AN_0 I'(0) - a$$

is negative or positive. Hence

$$I'(0) < \frac{a}{2AN_0} \tag{52}$$

is the condition for $\psi^* = 0$ to give a stable solution. When applied to the normal distribution this leads to inequality (16). In this case we cannot state that there may not be additional stable roots; however, if (52) does not hold then from the reasoning above and the fact that $F(\psi)$ is an odd function, as shown by (49), it follows that there is at least one stable positive root and also at least one stable negative root of equal absolute magnitude.

Now in addition to requiring that $n(\phi)$ and $p(\phi')$ be even functions, let us also require that they be unimodal, that is,

$$\begin{aligned} n(\phi_1) &\leq n(\phi_2) \quad \text{for} \quad |\phi_1| > |\phi_2| , \\ p(\phi_1') &\leq p(\phi_2') \quad \text{for} \quad |\phi_1'| > |\phi_2'| . \end{aligned} \tag{53}$$

With these additional restrictions on the distributions it is possible to show that the alternatives found for the normal distribution are the only possibilities; that is, either there is a single stable root at $\psi = 0$ or the root at $\psi = 0$ is unstable and there are two stable roots $\pm \psi^* \neq 0$. This results from the fact that under these conditions $I(\psi)$ is also unimodal as we now show. That $I'(\psi)$ is an even function is apparent from (50). From (47) we have

$$I'(\psi) = \int_{-\infty}^{\infty} p(\phi + \psi) n(\phi) \, d\phi , \tag{54}$$

so that

$$I'(\psi + h) = \int_{-\infty}^{\infty} p(\phi + \psi + h) n(\phi) \, d\phi = \int_{-\infty}^{\infty} p(\phi + \psi) n(\phi - h) \, d\phi , \tag{55}$$

and

$$I'(\psi - h) = \int_{-\infty}^{\infty} p(\phi + \psi) n(\phi + h) \, d\phi . \tag{56}$$

This gives

$$\begin{aligned} \Delta I'(\psi) &= I'(\psi + h) - I'(\psi - h) \\ &= \int_{-\infty}^{\infty} p(\phi + \psi) [n(\phi - h) - n(\phi + h)] \, d\phi . \end{aligned} \tag{57}$$

Since $I'(\psi)$ is even we also have

$$I'(\psi) = \int_{-\infty}^{\infty} p(\phi - \psi) n(\phi) \, d\phi , \tag{58}$$

and in the same way as (57) was obtained we find

$$\Delta I'(\psi) = \int_{-\infty}^{\infty} p(\phi - \psi)\,[n(\phi + h) - n(\phi - h)]\,d\phi\,. \tag{59}$$

Therefore we finally have

$$\Delta I'(\psi) = \tfrac{1}{2}\int_{-\infty}^{\infty} [p(\phi - \psi) - p(\phi + \psi)]\,[n(\phi + h) - n(\phi - h)]\,d\phi\,. \tag{60}$$

In this formula we take h and ψ to be positive; then for $\phi \geq 0$, (53) shows the second bracket to be ≤ 0 while the first bracket is ≥ 0, hence the product is ≤ 0. When $\phi \leq 0$ the signs of the two brackets are both reversed so the product is again ≤ 0. Hence $\Delta I'(\psi) \leq 0$ for $\psi > 0$, which completes the proof that $I'(\psi)$ is unimodal.

The general distribution with bias can also be treated. Suppose again that $n(\phi)$ and $p(\phi')$ are even functions, but that

$$N(\phi) = N_0 n(\phi - \phi_0)$$

with $\phi_0 > 0$. Then the value of I is

$$\begin{aligned} I_b(\psi) &= -\tfrac{1}{2} + \int_{-\infty}^{\infty} P_1(\phi + \psi)\,n(\phi - \phi_0)\,d\phi \\ &= -\tfrac{1}{2} + \int_{-\infty}^{\infty} P_1(\phi + \psi + \phi_0)\,n(\phi)\,d\phi = I(\psi + \phi_0), \end{aligned} \tag{61}$$

where $I(\psi)$ is the value of I for $\phi_0 = 0$. From the reasoning above there is always at least one stable positive root and there will be just one if $n(\phi)$ and $p(\phi')$ are unimodal. In the latter case there may be just one unstable negative root, $\psi = -\psi_1^* < -\phi_0$, and also just one stable negative root, $\psi = -\psi_2^* < -\psi_1^*$. Obviously for a stable negative root we must have $a/2AN_0 < I_b'(-\psi_1^*)$, and since $I_b'(-\psi_1^*) < I_b'(0) = I'(\phi_0)$ we have as a necessary, but not sufficient, condition for a stable negative root

$$\frac{a}{2AN_0} < I'(\phi_0)\,. \tag{62}$$

This work was aided by a grant from the Dr. Wallace C. and Clara A. Abbott Memorial Fund of The University of Chicago.

LITERATURE

Landahl, H. D. 1938. "A Contribution to the Mathematical Biophysics of Psychophysical Discrimination." *Psychometrika*, **3**, 107–125.

Rashevsky, N. 1948. *Mathematical Biophysics*. Revised Edition. Chicago: University of Chicago Press.

——— 1949. "Mathematical Biology of Social Behavior: III." *Bull. Math. Biophysics*, **11**, 255–271.

Rosser, J. Barkley. 1948. *Theory and Application of*

$$\int_0^z e^{-x^2}dx \quad \textit{and} \quad \int_0^z e^{-p^2y^2}dy \int_0^y e^{-x^2}dx\,,$$

p. 19. Brooklyn: Mapleton House.

BULLETIN OF
MATHEMATICAL BIOPHYSICS
VOLUME 12, 1950

A REMARK ON LANDAHL'S THEORY OF LEARNING

N. RASHEVSKY
COMMITTEE ON MATHEMATICAL BIOLOGY
THE UNIVERSITY OF CHICAGO

It is pointed out that the equations of Landahl's learning theory may be formally interpreted in terms of a different neural network than that considered originally by Landahl. The suggested interpretation is based on this author's theory of elimination of a wrong act through a delayed conditioned reflex which produces a reaction opposite to the wrong act.

H. D. Landahl's theory of learning (Landahl, 1941; Rashevsky, 1948) is based on the assumption of the neural network shown in Figure 1. The correct response R_c produces a change R_1 in the environment, which, acting itself as a stimulus, results in an increase of ε at the connection s_c. Similarly, the wrong response R_w via a change R_2 results in an increase of j at s_w.

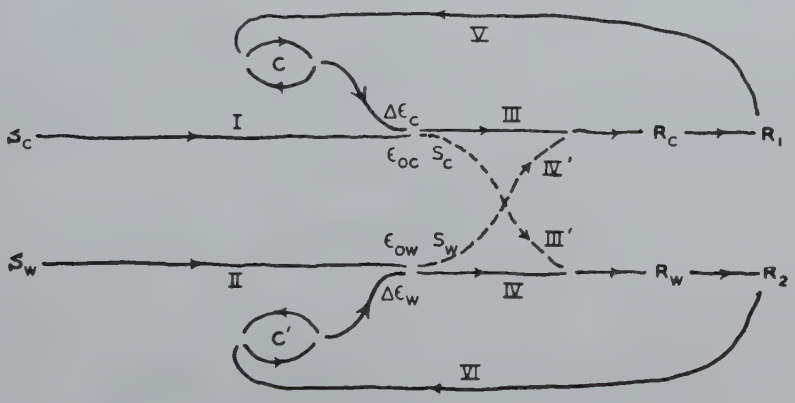

FIGURE 1

If interpreted too literally, the network of Figure 1 may lead to the misconception that the correct and wrong responses are already predetermined by the structure of the network, since the structure shown in Figure 1 permits only excitation of s_c and inhibition of s_w. Actually, s_c receives pathways from a number of cir-

cuits C: C_1, $C_2 \cdots C_n$. Some of those pathways are excitatory, some inhibitory. We shall call a circuit, C_i, which sends an excitatory pathway to s_c an excitatory circuit. A circuit C_i which sends an inhibitory pathway to s_c will be called an inhibitory circuit. All the circuits are stimulated by appropriate stimuli R_1: R_{11}, $R_{12} \cdots R_{in}$. The connection between R_c and R_1 is not a part of the network but part of the environment which is under the control of the experimenter. If R_c produces an R_{1i} which excites an excitatory circuit C_i, then R_c is a correct response. Otherwise, it is a wrong one. Thus R_c may be the entry into a given alley. But it is left to the experimenter to put either food (an R_{1i} which produces excitation at s_c via an excitatory circuit C_i) or a device for an electric shock (an R_k which produces inhibition at s_c via an inhibitory circuit C_k) at the end of the alley. The same holds regarding the connection s_w and the reaction R_w.

The purpose of this note is to show that formally the same results as in H. D. Landahl's theory are obtained by considering a previously discussed mechanism of elimination of a wrong act (Rashevsky, 1936; 1948). That theory is based on the idea that a wrong act results in a stimulus which produces a reaction opposite to the wrong reaction. Thus a dead end alley results in a return of the animal, that is, in a locomotion in the direction opposite to that of the original locomotion *into* the alley. This opposite reaction eventually develops as a delayed reflex to the original stimulus which produced the wrong reaction and, thus, reduces that wrong reaction in intensity. But this reduction in intensity is due not to an inhibition along the chain of pathway $s_c - R_c$, but to a purely physical mutual weakening of two opposite reactions.

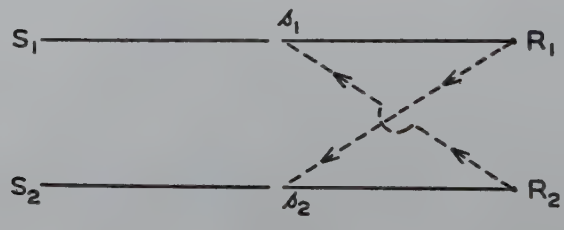

FIGURE 2

Now consider the network shown in Figure 2. Here the synapses s_1 and s_2 are inhibited by pathways which are stimulated not by S_1 and S_2, as in Landahl's well known network for reciprocal inhibition, but by the reactions R_1 and R_2. For stationary states we have the

same situation as in Landahl's standard case; namely, no reaction for $|S_1 - S_2| < h$, where h is the threshold of the system; reaction R_1 for $S_1 - S_2 > h$; and reaction R_2 for $S_2 - S_1 > h$. The application of the theory of random fluctuations leads to similar results as in Landahl's case. Now if one of the reactions—say, R_1—is the wrong one, it will be reduced by a process described above, and this will enhance the other reaction R_2, by reducing the inhibition at s_2.

The original theory of error elimination, mentioned above, does not provide for an increase in the correct response. This, however, may be easily obtained by a similar mechanism.

It may be noticed that learning will occur in Landahl's theory even without reinforcement of the correct response, but merely with a gradual inhibition of the wrong one. This amounts to putting the constant b (Rashevsky, 1948, p. 470) equal to zero. This cannot be done in the final equation [Rashevsky, 1948, p. 471, equation (7)] for reasons which the reader will readily see if he follows step by step the procedure outlined on p. 471 of *loc. cit.* However, if we put $b = 0$ from the beginning, we have, as in *loc. cit.*,

$$\frac{dw}{dn} = P_w = \tfrac{1}{2}\, e^{-k(\varepsilon_c - \varepsilon_w)}. \tag{1}$$

Using the same notations as in *loc. cit.* and putting, without loss of generality, $\varepsilon_{0c} - \varepsilon_{0w} = 0$, we find

$$\frac{dw}{dn} = \tfrac{1}{2}\, e^{-k\beta w} \tag{2}$$

or

$$e^{k\beta w} dw = \frac{dn}{2}. \tag{3}$$

Hence,

$$\frac{1}{k\beta}\, e^{k\beta w} = \frac{n}{2} + C, \tag{4}$$

where C is determined from the condition that for $n = 0$, $w = 0$. Hence $C = 1/k\beta$. Therefore, equation (4) gives

$$w = \frac{1}{k\beta} \log\left(1 + \frac{k\beta}{2} n\right). \tag{5}$$

The cumulative number of errors increases in this case indefinitely,

although dw/dn tends to zero. The ratio w/n decreases to zero, so that the task is learned better and better.

LITERATURE

Landahl, H. D. 1941. "Studies in the Mathematical Biophysics of Discrimination and Conditioning." *Bull. Math. Biophysics*, **3**, 13–26, 71–77.

Rashevsky, N. 1936. "Mathematical Theory of Delayed Reflexes in Connection with the Theory of Error Elimination." *Psychometrika*, **1**, 265–273.

——— 1948. *Mathematical Biophysics*. Rev. Ed. Chicago: University of Chicago Press.

BULLETIN OF
MATHEMATICAL BIOPHYSICS
BULLETIN OF

CONTRIBUTIONS TO THE MATHEMATICAL BIOPHYSICS OF THE CENTRAL NERVOUS SYSTEM WITH SPECIAL REFERENCE TO LEARNING

ALFONSO SHIMBEL
COMMITTEE ON MATHEMATICAL BIOLOGY
THE UNIVERSITY OF CHICAGO

A learning theory based on the lowering of thresholds of neurons under certain conditions is applied to two "random net" models. The first, a so-called "ganglion-brain" is characterized by completely random connections of all afferent tracts except certain ones which form the pathways for unconditioned responses. Certain expressions are derived which measure the learning potentiality of the ganglion — in particular, with respect to the number of responses which can be learned (conditioning potential) and the amount of interference between the learned responses (redundance potential).

The second model concerns the progressive refinement of a response. The efficiency of learning in this case is reflected in the eventual specificity of the response which, in turn, depends on the modification of the distribution of thresholds associated with the neurons governing the responses. Expressions are derived relating the initial distribution of thresholds, the relative effectiveness of the various responses, and certain other parameters to the final distribution of thresholds. For a particular choice of the effectiveness distribution of responses the progressive sharpening of the threshold curve (i.e., progressive specificity of response) is demonstrated. Some implications of the model with respect to the evolution of nervous systems are discussed.

It is now a rather commonly held hypothesis that the behavior of organisms possessing nervous systems (especially the more complex ones) will be greatly elucidated when the dynamics of neural tissue is better understood. Furthermore, it is held by many that one of the most fundamental aspects of neural dynamics depends in essence upon the topology of the connections (synapses) of neurons with one another, and the communication problems which such connections imply. These synapses and the patterns of neuron relations which they form (neural nets, as they are often called) are assumed to be of the greatest importance in all psychological phenomena. From this point of view the similarities between neural nets and the connections of vacuum tubes to one another in electronic computers, and the corresponding "intelligences" exhibited by these systems, are taken to be somewhat more than analogical.

N. Rashevsky, starting essentially with this point of view, was one of the earliest workers to give impetus to these notions and to systematically study their implications. This fundamental work is summarized in his book *Mathematical Biophysics.* Important advances of this approach made by H. D. Landahl and A. S. Householder are compiled in a monograph entitled *Mathematical Biophysics of the Central Nervous System,* wherein can also be found a summary of the ideas contained in a paper written by W. S. McCulloch and W. Pitts in 1943. The essential goal of that paper was an attempt to use the language of symbolic logic for the purpose of describing the behavior of neural nets of all kinds, where only certain "initial conditions" are known. Given information about the "state" of the neural net at some specified time, the method enables one to calculate all subsequent "states" of the net.

Although the attempt of McCulloch and Pitts represents an important step forward in neural net theory it has some serious shortcomings. For one thing, its definitive results are not generally applicable (at least in their present form) to neural nets containing cycles, i.e., to nets having closed pathways. Inasmuch as such networks are of general occurrence (Lorente de Nò, 1934), any mathematical technique purporting to describe neural dynamics cannot be considered adequate without subsuming such phenomena. Another shortcoming of the symbolic logic approach is the extremely laborious computations which must be made in order to analyze systems containing large numbers of neurons. Still another objection is the assumption that the system is somehow "locked" in phase, i.e., it is assumed that if any neuron in the system fires at some time $t = t_0$, then any other neuron in the system can fire either at exactly that same time or at some future time $t = t_0 + n\sigma$ where σ is taken as a constant and n is any positive integer. There has been some suggestion that this difficulty could be overcome by a slight modification of the mathematics. However, no such extension has appeared in the literature so that the point is still open to question.

The last and perhaps most serious shortcoming of the method of McCulloch-Pitts lies in the fact that it assumes an extremely detailed knowledge of the neural pattern under consideration. In fact, the method requires exact information concerning the relation of every neuron in the system to every other neuron. The gathering of such information concerning any one nervous system is a monumental task of incredible proportions and even if ever achieved would undoubtedly be inapplicable to any other system. Also, it is required that the threshold of every neuron in the system be known, the other

parameters being taken as equal for all neurons. Finally, given all of this information, the "state" of the system at some specified time $t = t_0$ is also required. This means that we must know of every neuron in the system whether or not it is firing at some time $t = t_0$. An impossible task.

These objections to the McCulloch-Pitts formulation, though they are indeed quite serious, cannot negate its importance as a formal approach to the subject.

In 1948 the author, together with A. Rapoport, published a paper in which was laid down a somewhat different approach to this problem. In this and subsequent papers an attempt was made to use the notions of probability in studying the structure and function of the nervous system. Many of the ideas developed in those papers thus parallel the development of cybernetics by N. Wiener and his associates and of the mathematical theory of communication by Shannon.

In essence, the problem which was encountered was that of inventing variables which could be used to characterize the state of a nervous system at any given time but which did not require any extremely specific knowledge concerning the disposition of the individual neurons of the system. Having defined such variables the next step was to show some of the necessary relations which they must bear to one another.

In order to mathematically symbolize the problem and to develop a formulation it was necessary to define certain components of nervous tissue in a highly schematic manner. These concepts are, of course, only approximations to actual nervous systems. Although the immediate application of such a formulation to the study of real nervous systems is tempered by the error incurred in its approximations and the complexity of the mathematics which it implies, the questions which it raises and its usefulness as a guide for further research are the ultimate measures of its value. These depend upon the future.

Let us briefly examine some of the ideas presented in this probabilistic approach. To begin with, those parameters which are associated with neurons as such (threshold, synaptic delay, refractory period, etc.) are assumed to vary from neuron to neuron according to some distribution function which, in general, would have to be determined empirically. The shapes of the distribution functions of these various parameters are assumed, in general, to vary from region to region in the neural net. The patterns resulting from the various neurons synapsing upon one another are characterized not by the dis-

position of each neuron in the system, but rather by certain functions of point pairs in space. These functions denote the *probability* that a neuron in a macroscopically small region about one of the points receives an axone from a neuron in a macroscopically small region surrounding the second point. The system is thereby viewed as an aggregate of neurons characterized by continuous (or stepwise) distributions of "local" properties together with certain "tendencies" of connection from region to region. A single neuron according to this picture, or even a small group of neurons, is not, in general, a determining factor in the "behavior" of the net. This seems more in accordance with the behavior of actual nervous systems wherein even widespread lesions often do not seriously impair the performance of the tissue as a whole.

Some mention should be made of the kinds of problems which these methods imply and attempt to deal with. They are essentially of three broad categories.

I. The Dynamics of Neuron Interaction.

This category of problems has already been briefly discussed in the foregoing. Several kinds of "sub-problems" have been approached from this point of view.

A. *Input-Output Relations in Nerve-Ganglion Systems.* Solutions for some special cases of this general problem have appeared in the literature (Shimbel, 1949; Rapoport, 1950a). Briefly the problem is as follows. Given an aggregate of, say, N neurons (see Figure 1) which is characterized by the kinds of parameters dis-

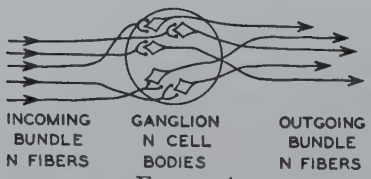

Figure 1

cussed in the foregoing and given also the frequency of action potentials occurring in the incoming bundle per axone per unit time (i.e., the input), determine the frequency of action potentials occurring in the outgoing bundle per axone per unit time (i.e., the output) as a function of the distributions of the various parameters in the ganglion.

B. *The Converse Problem.* Given a certain general relation

between the output and the input (not necessarily a specified functional relation) such as the existence of a maximum or the magnitude of the output as compared to the input at certain critical points, what relations (for example, inequalities) must be satisfied by the parameters for the desired relations to hold?

C. *Time Series.* If the input is considered as a time series (Shimbel, 1949) what distortion effects will appear as a result of a given distribution of parameters in the ganglion?

D. *The Converse Problem.* If the ganglion is considered as a transformation on a time series, what distributions of parameters will do a given job of transformation?

Having examined some of the properties of ganglia as to their effects on input-output relations we may perhaps be able to interpret more complex patterns of activity in terms of interactions between such "ganglion-elements."

In particular, the class of problems that has been designated as homeostatic problems may possibly be viewed in these terms.

II. "Hysteresis" in Neural Nets.

There are a number of well-defined problems which involve changes in the structure (and, consequently, the function) of neural networks as a result of the activity which occurs in them. Two questions arise in connection with such problems.

A. By what mechanisms does the activity of a neural net result in changes of its structure?

B. How, in general, do the changes in the structure of a neural net affect its dynamics?

The problem of learning evidently constitutes a special category of such problems, and perhaps its most important category. Other such phenomena, for example, fatigue and hyperexcitability may also be included.

III. The Evolution of Nervous Systems.

The problem of evolution of nervous systems as it appears from the point of view implicit in neural net theory is still rather poorly defined. Some ideas on the question, though still quite vague and speculative, have become themes for conversation among workers in the field, although very few have become sufficiently well-formulated for publication. (See, however, Rapoport, 1950b.) A few quali-

tative remarks concerning this question have also been included in this paper.

"Molecular" Theories of Learning

In many ways psychology and, in particular, learning theory bear a relationship to neural net theory which strongly suggests the relation which thermodynamics bears to kinetic theory. This comparison of psychology to thermodynamics and neural net theory to kinetic theory, if it serves no other purpose, at least expresses the hope that the "high order" concepts of psychology are essentially reducible to the language of communication theory. In what follows we will examine some of these reductionist attempts with respect to learning theory and briefly outline some possible alternatives.

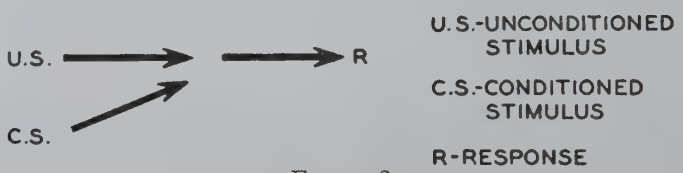

Figure 2

It has become standard practice to illustrate the principles of learning in "higher" animals by means of diagrams such as that shown in Figure 2. *U.S.* represents an unconditioned stimulus which invariably leads to a certain response *R*. *C.S.* represents a stimulus which before the learning process did not elicit the response *R*, but which thereafter is able to do so. The arrows, as used in the older psychological literature, chiefly served the purpose of symbolizing the above-mentioned relations among *U.S.*, *C.S.*, and *R*.

In the course of time, along with the development of neurophysiology and neural anatomy, the arrows in such diagrams began to take on a more definite "neural" meaning. Quite commonly now they are said to represent "neural pathways" or chains of neurons.

A great deal of anatomical and neurophysiological evidence has been accumulated which supports such a neural interpretation of "learning" diagrams. The details of such evidence are available in any modern textbook of neurophysiology and need not be presented here.

Let us start then with the assumption that a certain organism is capable of "detecting" a specific stimulus (call it *U.S.*), and that the "information" so detected is transmitted via a chain of neurons or a bundle of such chains (which, in general, may also have inter-

connections) to certain motor organs which then respond to this information (call this response *R*). Let us assume furthermore that this organism is capable of detecting a different stimulus (call it *C.S.*) which does not evoke the response given to *U.S.* Finally, we will assume that after these two stimuli are simultaneously or "nearly" simultaneously presented to the organism for a certain number of "trials," the stimulus *C.S.* alone becomes sufficient to evoke *R* and the information is again neurally transmitted. All of this is, of course, nothing more than a typical conditioning phenomenon.

How can the appearance of this "new pathway" be accounted for? Many attempts have been made to give this question a satisfactory answer. It has been suggested, for example, that the neurons of the *C.S.* bundle "grow" axones which synapse upon the neurons in the pathway from *U.S.* to *R*. This is the so-called theory of "neurobiotaxis." No such growth has ever been observed.

Another suggestion was that the necessary connections were already anatomically present and were absent only in a functional sense. The conditioning process, according to this point of view, serves merely to make these non-functional pathways operative.

If this is so, then we can ask: What is the mechanism by means of which such "non-functional" neural pathways can become operative through experience?

Reverberating Circuits.

N. Rashevsky (1938) described an interesting mechanism which accounts for at least the main features of this phenomenon. The simplest form of this mechanism is illustrated in Figure 3. Its

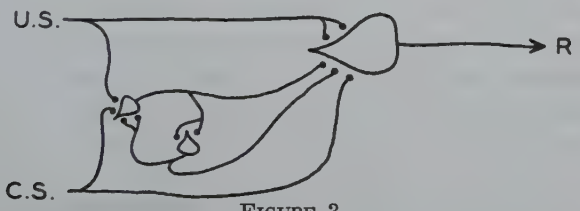

FIGURE 3

function is essentially as follows. Assume that the thresholds of all the neurons are equal to *two bulbs,* that is, at least two end bulbs terminating on a cell body must be "simultaneously" active in order to "fire" the neuron in question. It follows that each time *U.S.* occurs the large neuron leading to *R* will fire because the neuron stimulated by *U.S.* has two bulbs terminating upon it. Note that the axone of the neuron stimulated by *U.S.* also has a branch which leads to a

closed circuit of two mutually synapsed neurons but that this branch ends upon a neuron in the closed circuit with only *one* end bulb. Therefore, it is not capable of stimulating this neuron since all thresholds are assumed to be two bulbs. Note also that if *C.S.* occurs alone it will not fire either the neuron leading to *R* or the neuron in the closed circuit. Here again the insufficiency is due to the fact that the axone terminates on these neurons with only one bulb each, i.e., one bulb less than threshold.

Finally, observe that if *U.S.* and *C.S.* occur simultaneously (i.e., within the period of latent addition), then two things will happen.

1. The large neuron leading to *R* will be excited.

2. The effects of the neuron leading from *U.S.* will sum with the effect of that leading from *C.S.*, and as a result the closed circuit neuron will be excited.

Note that the closed circuit once stimulated will continue indefinitely. (Thus the name reverberating circuit.) Now if *C.S.* occurs alone, its effect on the neuron leading to *R* will sum with the effect of the branches of the closed circuit on the neuron leading to *R*, so that *C.S.* will now alone be a sufficient stimulus for *R*.

This, then, in its barest essentials is the principle of Rashevsky's model. Actually it has been much elaborated and refined and rather successfully applied to many aspects of learning theory. Further details can be found in Rashevsky's *Mathematical Biophysics*.

Synaptic "Resistance" Theories.

The so-called "synaptic resistance" theory is a possible alternative to Rashevsky's model. This theory is based on the conjecture that the frequent passage of impulses across a synapse tends to lower its threshold and thereby make the passage of an impulse "easier." Figure 4 illustrates how this idea can be adapted to the problem of conditioning.

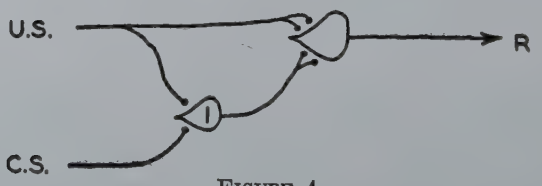

FIGURE 4

In this model, as before, all thresholds are assumed initially to be *two bulbs*. Clearly, then, *U.S.* is capable of eliciting *R*. However,

C.S. cannot excite the *R* neuron directly and has only one bulb terminating on the internuncial labeled *I* so that it has no evident effect on the system. However, note that if *C.S.* and *U.S.* are presented together, then not only will *R* fire but also *I* will be induced to fire by the combined action of the two stimuli. The theory implies that if *C.S.* and *U.S.* are presented together frequently enough (i.e., if *I* fires frequently enough), the threshold of *I* will be lowered so that eventually *C.S.* will become a sufficient stimulus for *R.*

A Simpler Model.

There are, of course, many other so-called "molecular" theories of learning. Any standard textbook of physiological psychology contains brief reviews of the more prominent ones. In what follows, however, we will be concerned with what may perhaps be viewed as a simpler version of the model just discussed. The first model to be considered is depicted in Figure 5.

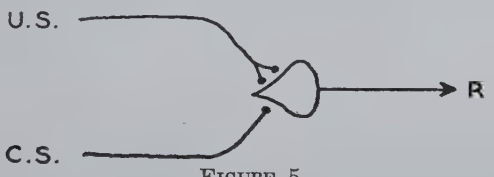

FIGURE 5

Here again, if we take the threshold of the neuron leading to *R* as *two bulbs,* then it is seen that *U.S.* is a sufficient stimulus for *R.* Similarly, it is clear that at the outset *C.S.* is inadequate. Now it is assumed that if *U.S.* and *C.S.* are frequently presented in temporal contiguity, eventually the threshold of the *R* neuron will fall below "normal" until finally *C.S.* will become a sufficient stimulus for *R.* It must be made clear that *U.S.* is assumed to deliver a superthreshold impulse, whereas *C.S.* is assumed to be *subthreshold.*

Our hypothesis explicitly requires that the threshold of a neuron will be lowered if a super-threshold and a subthreshold stimulus impinge upon it in sufficiently close temporal contiguity.

It should be admitted that all of the mechanisms suggested depend to a greater or lesser extent on *ad hoc* arguments. In fact, one "pays" for the simplification of the model by introducing new *ad hoc* hypotheses. This should, however, not in itself invalidate any attempt to develop a theory of learning from a particular chosen level.

An important question which arises at this point is: What is the mechanism responsible for this decrease in threshold? Variations

in threshold owing to various causes and under various conditions have, of course, been studied in some detail. Rapid and transient changes of threshold are consistently associated with active neural tissue. More lasting and more drastic threshold variations can readily be induced by a great variety of chemical agents. All of these observations tend at least to make plausible the assumption concerning threshold variation which is required for this molecular model of learning. One of the advantages of Rashevsky's model is that it does not require any such additional hypothesis but stems directly from the more usually considered properties of neurons. Be this as it may, we will assume in what follows that such threshold variations do occur and that, therefore, a learning theory based upon such phenomena is possible. It is suggested as an alternative to the Rashevsky picture because it seems at least superficially to require simpler neural circuits and, in general, fewer neurons.

The model as described so far is hardly more than a suggestion. It would appear natural at this point to examine in detail those aspects of neural metabolism most intimately associated with threshold variations and to attempt to formulate a physicochemical theory of such changes which may then be checked by experiment. The results so obtained may then more legitimately be compared to those derived from the Rashevsky model. In this paper, however, a rather different avenue of development will be attempted. We will be concerned in what follows with the general question: Given that learning is characterized by more or less permanent threshold changes such as those described above, how much *specificity* of growth is required in the ontogeny of a nervous system in order that it possess potentialities for learning?

The proportions and generality of such a question make any attempt to answer it in one fell swoop, or even to give it rigorous meaning, an almost hopeless task. It can, however, serve, so to speak, as a "definition" of an area of problems to be considered. Proceeding, then, in this spirit, we will pose and attempt to analyze some extremely simplified and hypothetical versions of the problem with the intention of using any results obtained thereby as an aid to the solution of progressively more realistic problems.

The Ganglion-Brain

Thus the first problem to be considered is the following. Assume that in some organism there exists a "central ganglion," i.e., a sort of primitive brain which is nothing more than a collection of, say, N neurons which synapse upon one another in a completely random

fashion. By randomness here we mean the following. In the course of the formation of the ganglion (i.e., when the neuroblasts are growing axones) there is no preferential synapsing. An axone growing out of a particular cell body is equally likely to synapse upon one other cell body as upon another. Operationally this may be described as follows. Let the neuroblasts be numbered in some way prior to the formation of the axones. After the axones have grown and the synapsing has occurred, let the number of each neuron be associated with the numbers of all the neurons to which it is directly connected. If no correlation can be established between the number of the neuron and the set of neurons to which it is connected, the connections will be said to be random. Assume, furthermore, that the afferent receptor tracts enter this aggregate and synapse randomly on the cell bodies within it. By "randomly" we again mean "without discernible order." In other words, the neurons upon which the efferent axones synapse are in no way distinguished (e.g. by their location) from each other. Finally, assume that some of the axones given off by the cells in the ganglion proceed peripherally to innervate the motor organs. (See Figure 6.)

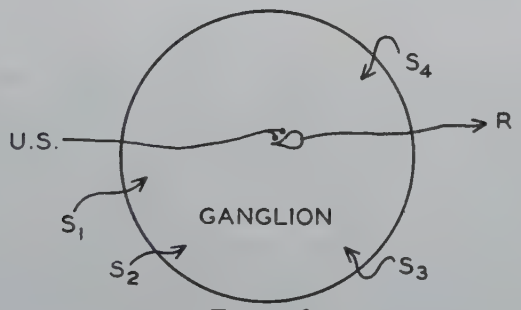

FIGURE 6

Now suppose we examine the neural pathway which is followed by an unconditioned stimulus, from its detection at some receptor to the motor organs which eventually respond to it. Such a stimulus will follow a chain or bundle of afferent fibers, which lead into the central ganglion, wherein, through a number of internuncials, it will eventually arrive at the efferent pathways which lead out of the ganglion toward the organs which give the response. For simplicity, let us assume that the afferent and efferent pathways are not bundles of neural chains, but simply one such chain of serially connected neurons joined by a single internuncial within the ganglion.

Let us now consider any other receptor pathway (again a simple chain) leading into the ganglion and there synapsing randomly

upon one of the ganglion neurons. Given that the learning mechanism is a threshold phenomenon as described above, can the motor organ which unconditionally responds to the stimulation of the first pathway be conditioned to respond to the stimulation of the second? An unambiguous answer to this question can be given only if we know whether or not the second receptor chain eventually (i.e., through a series of internuncials) connects to at least one of the neurons in the first chain. Since the connections in the ganglion are assumed to be random, this question cannot be answered by a *yes* or *no* but only by a probability statement. Rephrasing the question, we ask: Given any two neurons in such a randomly connected ganglion, what is the probability that the first neuron will, through any number of internuncials, be connected to the second? This problem has been solved for a special case by A. Rapoport (1948). It is supposed there that every neuron in the ganglion has exactly *one* axone which synapses with equal probability on any one of the neurons in the ganglion. It is shown that under these assumptions the probability C that an arbitrarily selected neuron will be a member of a cycle is given by the equation

$$C = \sqrt{\pi/2N}, \tag{1}$$

where N is the number of neurons in the ganglion. This is exactly the probability we are seeking, since the probability of starting the chain at one neuron and ending it on the *same* neuron is the same as starting the chain at one neuron and ending it on any other *specified* neuron.

To be sure, this probability is small for large N. If, however, we suppose that not one but several axones emanate from each neuron in the ganglion, then the probability of an arbitrary pair being connected rises rapidly with the number of axones.

To view it in another way we can consider the first neuron of the arbitrary pair as an analogue to a person infected with a contagious disease living in a closed and healthy population. The total number of persons who ultimately succumb to the disease in the resulting epidemic will presently be shown as analogous to the total number of neurons ultimately connected to the first. The ratio, then, of this number to the total number of neurons in the ganglion will be the probability that there exists a path of internuncials leading from the first to the second.

Analogy between the Contagion Equation and Probability of Path Existence.

Suppose first that the disease is not lethal; then every infected individual continues to live and infect others. The rate of infection in a population of n individuals will then be governed by the following differential equation

$$\frac{dx}{dt} = kx(n - x), \tag{2}$$

where x is the total number of infected individuals (and, therefore, $(n—x)$ is the total number of healthy individuals) and k is a time constant governing the rate of spread of the disease. It is easy to see that ultimately every individual in the population will become infected.

In fact, the number of individuals which has become infected is given by the well-known solution of the "logistic equation" (2)

$$x(t) = \frac{nAe^{nkt}}{1 + Ae^{nkt}},$$

where $A = x(0)/[n - x(0)]$.

Translating this situation into terms of neurons, axones, and connections, this would mean that a single neuron begins the "infection" by giving off axones at a constant rate and each neuron which receives one of these axones begins, in turn, to grow axones at the same rate affecting still other neurons in the same way. Clearly, then, any neuron which at any time is thereby brought into this system will be accessible through a pathway leading from the instigating neuron. As in the case of contagion, every neuron will ultimately be brought into the system. Note, however, that the actual situation in our ganglion is somewhat different. The difference lies in the fact that in our ganglion each neuron gives off only a limited number of axones, whereas the contagion equation implies that each neuron continues to give off axones indefinitely. This difficulty can be remedied by a proper modification of the contagion equation. This modification can be considered, in the language of epidemic, as the introduction of a death rate with a constant time lag between infection and death. The death of an individual is then analogous to the exhaustion of any one neuron's "supply" of axones. Just as the former individual stops infecting others, so does the latter stop making new connections.

We introduce the following variables in the modified contagion equation:

$x(t) \equiv$ Number of individuals sick or dead at time t,
$y(t) \equiv$ Number of individuals dead at time t,
$n \equiv$ Total number of individuals in original population,
$\tau \equiv$ Time lag between infection and death,
$k \equiv$ Time constant.

In this notation we then have

$$dx/dt = k(x - y)(n - x), \tag{3}$$

$$y = \begin{cases} 0 \text{ for } t \leq \tau, \\ x - \int_{t-\tau}^{t} x'dt \quad \text{for} \quad t > \tau, \end{cases} \tag{4}$$

whence we obtain $y(t) = x(t - \tau)$, for $t > \tau$. Substituting this into (3), we obtain

$$\frac{dx}{dt} = k[x(t) - x(t - \tau)](n - x). \tag{5}$$

The problem now reduces to the solution of the following system:

$$\begin{aligned} \frac{dx}{dt} &= kx(n - x), \quad 0 \leq t < \tau, \\ \frac{dx}{dt} &= k[x(t) - x(t - \tau)](n - x), \; t > \tau. \end{aligned} \tag{6}$$

The system (6) can also be expressed by equation (5) alone with the initial condition on the function $x(t)$, namely

$$x(z) = 0 \quad \text{for} \quad z < 0.$$

The solution of this system behaves for $t < \tau$ in exactly the same way as the solution of the ordinary contagion equation (2). However, increasing τ corresponds in the neuron-axone language to a large number of axones per neuron, and so, as one would expect, for a sufficiently large number of such axones, the probability of any specified connection approaches unity.

Learning as Simple Conditioning.

Returning now to the model depicted in Figure 6, it should be

pointed out that the assurance of a ganglion with sufficiently dense connections does not necessarily imply an adequate learning mechanism. Although in such a model any response can be conditioned to almost any stimulus, a serious difficulty arises in that the lowering of threshold concomitant with a particular learning experience will, in general, involve many other afferent pathways. In other words, a response may be conditioned to many different stimuli without ever having been associated with them. Such conditioning we shall call "redundant conditioning."

To be sure, the phenomena of abstraction and generalization involve just such processes and can conceivably be explained by such "redundant conditioning." However, on a different level the "abstractions" are the "stimuli" to be differentiated. Here again one may argue that a higher generalization process may take place, perhaps again by the mechanism of redundant conditioning. It is clear, however, that for any given level of abstraction, an effective nervous system will provide for a certain optimum specificity. Thus, for a given level of abstraction, an "optimum net" represents some sort of balance between spread of response and specificity. Whatever that optimum may be, it is clear that "total redundancy," i.e. where all conditionable pathways are conditioned together, is hardly desirable. In such a situation, "everything is everything else" for the organism and discrimination is impossible.

This difficulty can be overcome by a generalization of the model. Assume now that the unconditioned stimulus-response pathway consists not of a single neural chain but makes up a bundle of s such parallel chains, each of which has one neuron in the ganglion. This model is illustrated in Figure 7.

The other afferent fibers constituting the pathways for other stimuli also arrive as tracts rather than single neural chains. These

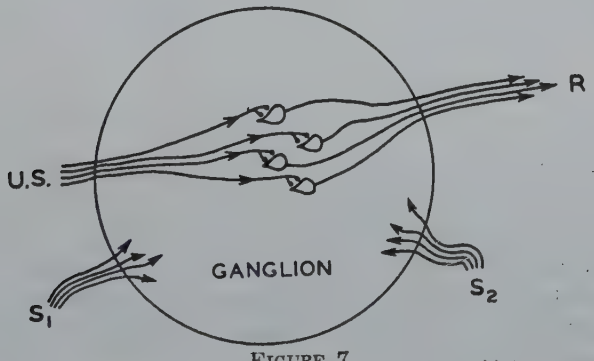

FIGURE 7

then synapse randomly on the cell bodies of the ganglion. At this point a somewhat different concept of threshold is introduced. In order for a given response to occur in the motor organ we require that a certain minimal number, $h < s$, of the fibers (single neural chains) innervating it must be active. Clearly this condition is always satisfied for a sufficiently strong stimulus traversing the unconditioned pathway. However, it is possible for this response to be conditioned to any other stimulus only if at least h distinct efferent pathways are accessible to the stimulus in question. As has been shown above, this condition will prevail when the interconnections of the ganglion neurons are sufficiently dense. Here again, since we assume that the interconnections in the ganglion (other than the unconditioned pathway) are random, we can speak only of the probability that for any given case the requirement for the possibility of conditioning be met. This probability can be expressed in terms of s, h and p, the probability of a single efferent pathway being accessible to the afferent tract. In fact, this is simply the expression for the probability of at least h successes in s independent trials, each with a constant probability p. The probability C_h of exactly h successes in s trials is given by (Uspensky, 1937)

$$C_h = \frac{1}{\sqrt{2\pi spq}} e^{-\frac{t^2}{2}} \left[1 + \frac{(q-p)(t^3 - 3t)}{6\sqrt{spq}} \right] + \Delta, \qquad (7)$$

where $t = h - sp/\sqrt{spq}$ and

$$|\Delta| < \frac{0.15 + .25\,|p-q|}{(spq)^{3/2}} + e^{-\frac{3}{2}\sqrt{spq}},$$

provided $spq \geq 25$.

Thus the probability C of *at least* h "successes" is given by the expression

$$C(h, s, p) = \sum_{j=h}^{s} C_j. \qquad (8)$$

If h, s, and p are considered as parameters of the ganglion, then $C(h, s, p)$ is also a parameter of the ganglion. We shall refer to it as the "conditioning potential." It is a measure of the number of pathways of the ganglion which can be conditioned to the unconditioned stimulus-response pathway. Evidently, then, organisms with a high conditioning potential will have a greater potential repertoire of learned responses.

However, as we have seen, the usefulness of a learning mechanism depends not only upon the high probability of effective connections but also upon the non-interference of the various afferent tracts with one another. In our present model it may appear that as in the previous case (single chain afferents), the high accessibility of efferent pathways to all stimuli also implies the high probability of "redundant" conditioning. In order to illustrate how redundant conditioning can be avoided by our generalized model while high accessibility of efferent pathways to the afferent tracts is still maintained, we will refer to Figure 8. Note that the pathways from S_1 and S_2 to R are drawn as *direct* connections. This need not be the case. In general, these connections occur through internuncials.

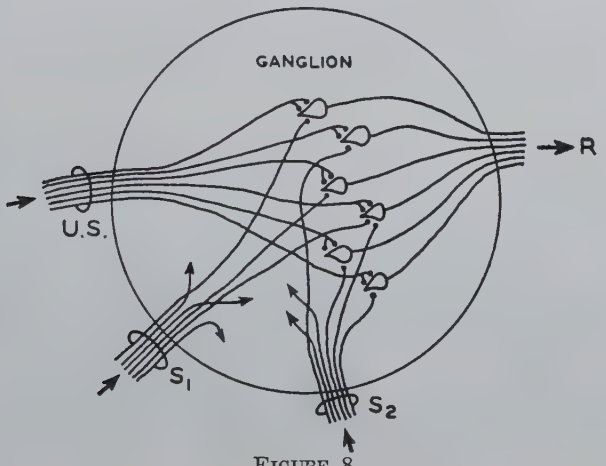

FIGURE 8

Here, as usual, the afferent fibers which carry the unconditioned stimulus *U.S.* synapse on an equal number of cell bodies in the ganglion which then send axones (efferent) to the outgoing bundle which stimulates R. For purposes of illustration, let us arbitrarily take $h = 3$ and $s = 6$. A glance at Figure 8 shows that both S_1 and S_2 satisfy the conditions for potential conditioning to R; i.e., both S_1 and S_2 have $h = 3$ common paths with *U.S.* Nevertheless the conditioning of R to S_1 does not imply the conditioning of R to S_2. This is so because less than h of the paths which S_1 and S_2 each have in common with *U.S.* are overlapping paths. In other words, those neurons whose thresholds are lowered as a result of conditioning R to S_1 are not all the same neurons whose thresholds must be lowered in order to condition R to S_2.

Of course, as h approaches s the probability of "redundant"

learning increases. Also, as p approaches unity the probability of redundant learning increases.

Now we seek an expression for the probability of redundant conditioning in a net such as that illustrated in Figure 8. That is to say, suppose that two afferent tracts S_1 and S_2 each have a sufficient number of paths in common with the ganglionic internuncials of *U.S.* leading to R so that R can be conditioned to either of them. Suppose further that the numbers of common paths S_1 and S_2 have with *U.S.* respectively are a and b, $a \geq h$, $b \geq h$. We ask: What is the probability that h or more of these pathways are common to S_1 and S_2?

This probability can be computed in terms of the solution to the following mathematically equivalent problem. Suppose an urn contains m white and n black balls. If one removes λ balls from the urn, what is the probability, P, that μ of the λ balls will be white and ν will be black?

The analogy between this problem and ours can be seen as follows. The number, a, of common paths between S_1 and *U.S.* corresponds to the m white balls in the urn. The $(s - a)$ remaining paths of *U.S.* correspond to the n black balls. The number, b, of common paths between S_2 and *U.S.* corresponds to λ, the total number of balls removed from the urn. We now ask: How many of the b paths common to S_2 and *U.S.* are "white," i.e., also common to S_2 and *U.S.*, and how many are "black," i.e., *not* common to S_2 and *U.S.*?

The answer to this problem of white and black balls was given by A. Markoff (1913). It is

$$P = \frac{(\mu+\nu)!\, m(m-1)\cdots(m-\mu+1)\, n(n-1)\cdots(n-\nu+1)}{\mu!\nu!\,(m+n)\,(m+n-1)\cdots(m+n-\mu-\nu+1)}. \qquad (9)$$

Translating this expression into our notation we obtain the following for the probability, $P(a,b,r)$, of r "redundant" connections between two afferent tracts of respective numbers of "conditionable" paths, a and b,

$$P(a,b,r) = \frac{b!\, a(a-1)\cdots(a-r+1)\,(s-a)\,(s-a-1)\,(s-a-b+r+1)}{r!\,(b-r)!\, s(s-1)\cdots(s-b+1)}. \qquad (10)$$

The probability that the redundancy r exceeds threshold h is given by the expression

$$P(a,b) = \sum_{r=h}^{c} P(a,b,r), \qquad (11)$$

where

$$c = \min(a,b). \tag{12}$$

From the quantity $P(a, b)$ and the components C_j of the conditioning potential it is possible to write an expression for the "redundancy potential" of the ganglion. This is given by

$$\mathcal{R}(s,h,p) = \sum_{j=h}^{s} \sum_{i=h}^{s} C_i C_j P(i,j). \tag{13}$$

The redundancy potential is, therefore, simply the sum of the P's properly weighted by the C's which denote the probability of their occurrence.

Thus C measures the number of things which can be learned, while $\mathcal{R}$ measures the interference between the various learned responses.

Implications of the Model.

So far it has been shown that a collection of neurons having only certain assumed properties of threshold variation and possessing only a minimum of specificity in its connections (the unconditioned tracts are assumed to lead directly to motor responses) can be connected in a completely random fashion and nevertheless possess simple properties of learning.

We will now examine some of the implications of this model which have not yet been mentioned and some of the problems which they suggest.

The Optimal Net.

As was mentioned in the foregoing [see expression (8)], the probability that a given afferent tract has enough paths in common with any specified unconditioned pathway (what we have called the conditioning potential) depends upon three variables, namely, p, s, and h. As the probability p, that any one chain of the unconditioned tract is linked with the afferent tract, rises, so also the conditioning potential C rises. But as p approaches unity, a and b approach s. Under these circumstances the probability that the redundancy R between any two afferent tracts exceeds h approaches unity.

In other words, if we wish to guarantee that every unconditioned tract has a sufficient number of paths in common with every other afferent tract, redundancy of connections exceeding threshold will become a certainty. In the extreme case we would have an animal

which could associate any two stimuli with the same response but once having done so for any two stimuli would automatically have associated every other possible stimulus with that same response. An animal possessing a "brain" near this extreme would readily learn but would equally readily "confuse" what it had learned. It would quickly reach a "saturation" point where it could, literally, not distinguish any more new "information."

It is seen then that conditioning potential and redundancy are in a sense antagonistic to one another. It may perhaps be possible to find a "natural" definition for some kind of optimum balance between these two tendencies. This would lead to a well-posed mathematical problem of determining the relations between the variables which must prevail in order for the "ganglion-brain" to achieve this optimum balance.

A possible suggestion is the following. Since C measures the probability that a given stimulus has adequate connections to a given response, the quantity $1 - C$ is the probability of failure. The quantity $\beta(1 - C)$ would be a measure of the "loss" which is incurred by the organism because of this inadequacy. The constant β would of course depend upon the particular environment which the organism inhabits. If now we assume that maximum specificity of response is optimal then the quantity $\gamma\mathcal{R}$ will also be a measure of a certain "loss" to the organism. The constant γ is again determined by the environment. The total "loss" to the organism would then be given by the expression

$$L = \beta(1 - C) + \gamma\mathcal{R}. \tag{14}$$

The construction of the "optimum net" would then involve the minimization of expression (14). This in turn imposes conditions on the parameters of the net upon which both C and $\mathcal{R}$ ultimately depend.

This definition of "optimum net" is obviously a very special one, since it assumes that maximum specificity is the most desirable. More generally this need not be the case. The general theory of optimal nets leads to a separate set of problems.

Specific Learning Abilities.

Suppose we would modify the connections in the ganglion so that instead of being completely random, certain regions would have axones which *tend* to grow in some given direction. We would evidently find that such an animal's ability to associate certain stimuli

with a given response would be exceptionally high as compared with other stimuli. If a population of such animals possessing different "biases" in the direction of axones in their "ganglion-brains" were tested for their ability to learn certain simple patterns we would find that certain members of the population would be especially "intelligent" with respect to any specified pattern. On the other hand, if we compared the learning ability of the "intelligent-for-pattern no. 1" group with that of the "stupid-for-pattern no. 1" group with respect to pattern no. 2, we would have no a priori reason, judging from our model, to assume that the "intelligent" group would retain its superiority for the second pattern. This is reminiscent of R. C. Tryon's work (1931) in the selection of rats which were exceptionally capable of solving certain mazes. When tested with new mazes the exceptional rats showed only ordinary ability.

Relation to Evolution.

Suppose that some population of animals possessing such ganglion-brains with various hereditary growth "biases" is living in a more or less constant environment. In general, we would expect that the ability to make certain simple "associations" would increase the likelihood of its survival. Furthermore, any mutants which enhance such growth "tendencies" without significantly changing the other features of the species would also be selected for. Such a selection process, it would seem, should eventually lead to some rather specific neural orientation especially adapted to the enhancement of the "valuable" association. In the process of such evolution we might expect that eventually the members of the population would have such a great density of "right" connections that learning might be simply a matter of a few trials. At this point slight hereditary changes in the neural thresholds would make the response "instinctive."

These considerations may indicate the way to answer the perplexing question concerning the evolution of complex "instinctive" behavior patterns. It has always been difficult to imagine that such patterns arose through gradual synthesis of simpler "units of behavior," because the simpler units by themselves did not appear to have any survival value. It was hard to conceive that the mud wasp had the habit of flying about with a ball of mud for a million years until it finally decided to use it in house building. Why should a habit of doubtful survival value have been retained so long?

The implications of the ganglion-brain random net model view the evolution of instinctive behavior as derived from more and more *easily learned* patterns, which become gradually more intimately

linked through the evolution of biased connections.

Organization versus Chaos.

Notice that in this primitive organism which evolved a simple innate behavior pattern, a portion of its *random* neural tissue has been "sacrificed." As more and more such innate patterns evolved, progressively more random tissue would become organized. The newly acquired innate behavior patterns are, of course, assumed to be an "advantage" to the species, but, on the other hand, the ability to learn new responses (which ability depends, according to our model, upon *random* tissue) is presumably also of importance. We might expect then that in the course of its evolution the species would "attempt to compensate" for this loss by increasing the size of its "ganglion-brain" and thereby replenish some of its random tissue. Notice also that the geometry of such development would imply that the more highly channelized and older circuits would find themselves more and more deeply imbedded in the less specialized tissue. There is some evidence to indicate that this is actually the case.

Progressive Refinement of Response in Learning

The kind of learning so far discussed is of a rather simple nature in which a learned response is either categorically given or not given. In reality most responses are graded. Learning is usually not a question of *yes* or *no,* but, rather, how much. In any coordinated activity an optimal response to a given stimulus involves not merely the combined activity of a group of motor organs but, in general, requires that each element in the combination respond with a certain "optimal" intensity. How can such adjustments of graded responses be linked with the learning processes?

We will describe an exceedingly simple model which may account for such phenomena. The model lends itself to rather obvious generalizations. For greater clarity the simple version will be discussed in detail. The possible generalizations will then be indicated.

In the model which we are about to consider the grading of a response will be represented by a spatial coordinate. This procedure will not be unfamiliar to those acquainted with the "spatializing" of graded responses in physiological psychology as, for example, in the theories of hearing. In our model, therefore, the "narrowing down" of a response toward an "optimum" will be represented as a progressive localization. This "localization" is actually an artifact of our

representation and need not imply geographic localization in the nervous system. The model is illustrated by Figure 9.

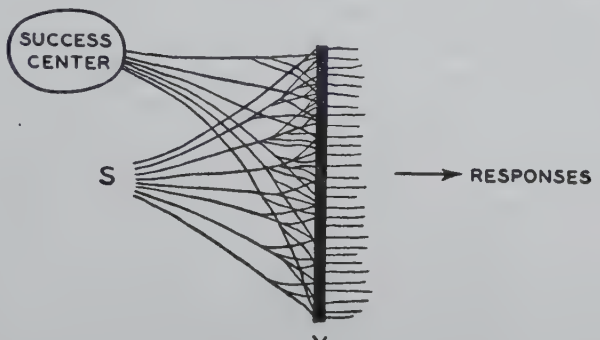

FIGURE 9

Let S represent some stimulus (for simplicity of discussion let us consider it to be a noxious stimulus) whose afferents innervate an aggregate of neurons which in Figure 9 is represented by a coordinate x. The instantaneous thresholds h_i of the neurons in the array are assumed to fluctuate about some *mean value* $\bar{h}$. This quantity is a function both of x and, as will be shown below, *implicitly* of time. Thus

$$\bar{h} = \bar{h}(x,t). \tag{15}$$

Future references to the "threshold" will imply this *mean* value $\bar{h}$. We shall refer to $\bar{h}(x,0)$ as the initial distribution of thresholds $\bar{h}_0(x)$. For any given subthreshold stimulus intensity we should expect that the instantaneous threshold h_i will, owing to its fluctuations, occasionally fall sufficiently low to permit the neurons associated with it to fire. However, if $\bar{h}$ is too large at certain points on the linear array and the amplitude of the fluctuations never exceeds a certain maximal amplitude, then the neurons associated with these points would, for certain stimulus levels, be effectively non-functional.

Now to proceed with the description of our model, we will assume that the firing of any neuron in the array results in motor activity which manipulates the environment in such a way as to increase or decrease the noxious stimulus S. This effect E on S is not the same for all neurons in the array but, in fact, is a function of the "position," x, of the neuron in the array: $E = E(x)$.

An essential feature of the model is the so-called success center.

Here again the word *center* is used in a functional rather than anatomical sense. A certain level of activity is assumed always to be present in the center. The neurons in it are innervated by tracts leading from the receptors of the noxious stimulus S in such a way that its activity is *inhibited* in proportion to the intensity of S. Thus the decrease of the intensity of S tends to increase the level of activity in the center. This implies that the neurons in the linear array (by stimulating motor activity which in turn manipulates the environment and thus affects S) thereby have a differential effect upon the activity in the success center. This differential effect may be represented by a function of "position" x, which may, in fact, be taken to be $E(x)$. The activity of the success center lowers the thresholds of the linear array in accordance with the previously suggested mechanism, i.e., the amount of lowering of the threshold of a particular neuron depends not only upon the activity of the success center but also on the activity of the neuron.

We will suppose that $\bar{h}(x)$ is lowered by an amount proportional to the number of times, N, the neuron at x has fired. (To be exact, N refers to the *expected* number.) In order to express N as a function of the parameters of the model it will be necessary to introduce some notation. Let $\eta(\xi)$ denote the relative frequency distribution of the instantaneous threshold h per unit time. The probability per unit time that the instantaneous threshold h_i is below the stimulus level S is then given by the expression

$$\int_0^S \eta(\xi)\, d\xi . \tag{16}$$

The expected number of times N that the neuron in question has fired up to time t will, therefore, be given by the relation

$$N = \int_0^t \int_0^S \eta(\xi)\, d\xi d_\tau . \tag{17}$$

To express the lowering of the initial threshold of a neuron by the activity of the success center we have the equation

$$\bar{h} = \bar{h}_0 - \alpha EN , \tag{18}$$

where α is a constant and where $\bar{h}$ is taken to be identically zero for all $\alpha EN > \bar{h}_0$. This latter condition is necessitated by the consideration that negative $\bar{h}$ is physically meaningless. In other words, we suppose that the decrease in threshold is linear and proportional to

E, the effectiveness of the activity resulting from the firing of the neuron with respect to decreasing the intensity of the noxious stimulus, S, (and thus increasing the activity of the success center). It is also proportional to N, the number of times the neuron has fired (and thus had its threshold lowered by the activity of the success center).

Introducing the expression for N [equation (17)] into equation (18) we obtain

$$\overline{h} = \overline{h}_0 - \alpha E \int_0^t \int_0^s \eta(\xi)\, d\xi d\tau . \tag{19}$$

If the distributions $\overline{h}_0(x)$, $E(x)$, and $\eta(\xi)$ were known we could theoretically solve for $\overline{h}$ in terms of x and t. The narrowing down of the response to the optimal one would then consist in the "sharpening" of $\overline{h}(x,t)$, considered as a function of x, as time progresses. In other words, learning would be reflected in the appearance of a minimum in the graph of $\overline{h}(x)$ and in the progressively narrowing "invagination" of the graph around that minimum. The appearance of more than one minimum would imply that alternative optimal responses to the stimulus were possible.

The sharpness of the graph at the minimum will imply that a certain neuron of the linear array is much more likely to fire as a response to the stimulus than any other neuron in the array. The net will have learned to respond in a specific way to the stimulus.

In order to illustrate this phenomenon we can assume that the distributions appearing in equation (19) are of very simple form. In particular let us put $\overline{h}_0 =$ constant. Furthermore, let $\eta(\xi)$ be a rectangular distribution of width $2A$. This simply says that all possible fluctuations occur with equal frequency.

Note that according to these assumptions it may be that for some of the neurons in the array we would have $\overline{h} > S + A$. This would mean that such neurons would never be active. If the stimulus level were raised, however, or if the range of fluctuation A were increased, then such neurons would again become functional. For the sake of definiteness we will exclude such neurons from our consideration, i.e. we will assume in what follows that $\overline{h}$ is always smaller than $S + A$. Note also that if the mean value of the threshold of some neuron is lower than the stimulus level S, such a neuron would be firing almost continuously whenever S is acting. This may be considered analogous to an instinctive or reflexive response to

a specific stimulus. We will also exclude these cases from our considerations and take our analysis to be meaningful only so long as $\bar{h} > S$.

Using these assumptions then, after a suitable choice of time units, equation (17) becomes

$$N = \int_0^t [S + A - \bar{h}(x,\tau)]\, d\tau . \tag{20}$$

Now by letting $I = S + A$ and substituting the expression for N into equation (19) we obtain

$$\bar{h} = \bar{h}_0 - \alpha E \int_0^t [I - \bar{h}]\, d\tau ; \tag{21}$$

whence by differentiation

$$\frac{d\bar{h}}{dt} = -\alpha E\,(I - \bar{h}) , \tag{22}$$

so that

$$\frac{d\bar{h}}{(\bar{h} - I)} = \alpha E dt. \tag{23}$$

This leads to the solution

$$\log C(\bar{h} - I) = \alpha E t , \tag{24}$$

where C is the constant of integration. The evaluation of C by setting $t = 0$ yields

$$C = (\bar{h}_0 - I)^{-1}; \tag{25}$$

therefore

$$\frac{\bar{h} - I}{\bar{h}_0 - I} = e^{\alpha E t}, \tag{26}$$

whence

$$\bar{h}(x,t) = I - [I - \bar{h}_0(x)] \exp [\alpha E(x) t] , \qquad I > \bar{h} > I - A . \tag{27}$$

Now in order to obtain $\bar{h}$ as an explicit function of x and t we must stipulate the form of the distribution $E(x)$. Let us choose the origin of the linear array to be at the center, and the effectiveness, E, to be of the form $E(x) = 1/(1 + x^2)$. We are interested in the

graph of $\bar{h}(x, t)$ as it changes with time. In particular, does a minimum appear, and does the second derivative at that minimum grow? These criteria are not sufficient to insure the "sharpness" of the graph at the minimum. We can, however, examine other criteria, for example, the coming together of the inflexion points on both sides of the minimum and the value of the first derivative at the inflexion points. Again we may measure the sharpness of the minimum by the standard deviation of the function $\bar{h}(x, t)$ considered as a distribution.

The origin of the linear array is taken to be at the center, and the effectiveness, E, is greatest there. Furthermore, let I and $\bar{h}_0$ be constant throughout the array. We are interested in the graph of $\bar{h}(x, t)$ as it changes with time. In particular, does a minimum appear, and does the second derivative at that minimum grow? These criteria are not sufficient to insure the sharpness of the graph at the minimum. We can, however, examine other criteria, for example, the coming together of the inflexion points on both sides of the minimum and the value of the first derivative at the inflexion points. We note that

$$E'(x) = -\frac{2x}{(1+x^2)^2}; \tag{28}$$

$$E''(x) = \frac{-2(1+x^2)^2 + 8x(1+x^2)}{(1+x^2)^4} = \frac{6x^2-2}{(1+x^2)^3}. \tag{29}$$

Both derivatives (28) and (29) are non-positive in the neighborhood of $x = 0$. Furthermore,

$$\frac{\partial \bar{h}}{\partial x} = -(I - \bar{h}_0)\alpha t e^{\alpha E t} E'(x). \tag{30}$$

This vanishes only where $E'(x) = 0$, that is, at $x = 0$. For the second derivative, we have

$$\frac{\partial^2 \bar{h}}{\partial x^2} = -(I - \bar{h}_0)\alpha^2 t^2 e^{\alpha E t}(E')^2 - (I - \bar{h}_0)\alpha t e^{\alpha E t} E''. \tag{31}$$

This is positive at $x = 0$. Thus the existence of a unique minimum is assured. From equation (31), we see that $\partial^2 \bar{h}/\partial x^2$ does increase with time at the minimum; in fact, it increases rather rapidly since it is proportional to $te^{\alpha E t}$.

To obtain the behavior of the inflexion points, we set $\partial^2 \bar{h}/\partial x^2 = 0$, whence

$$-at(E')^2 - E'' = \frac{at4x^2}{(1+x^2)^4} + \frac{6x^2-2}{(1+x^2)^3} = 0, \tag{32}$$

which, after simplification and rearrangement, yields

$$6x^4 + 4x^2(at+1) - 2 = 0. \tag{33}$$

The roots of equation (33) give the positions of the inflexion points in the graph of $\bar{h}(x)$ for each t. By the symmetry of $h(x)$, they are equidistant from the origin, and the distance between them is measured simply by the coordinate of either one of them. But the solution of (33) for x^2 gives for the significant (positive) root

$$(x^*)^2 = \tfrac{1}{3}\left[-at - 1 + \sqrt{(at+1)^2+3}\right]. \tag{34}$$

We seek the behavior of $(x^*)^2$ as time increases. Its derivative with respect to time is given by

$$\frac{d[(x^*)^2]}{dt} = \frac{a}{3}\left[\frac{2}{\sqrt{(at+1)^2+3}} - 1\right]. \tag{35}$$

We wish to know whether the right side of (35) is negative, i.e., whether the points of inflexion are approaching each other. This will be so if

$$\frac{2}{\sqrt{(at+1)^2+3}} < 1, \tag{36}$$

$$2 < \sqrt{(at+1)^2+3}, \tag{37}$$

or

$$4 < a^2t^2 + 2at + 4. \tag{38}$$

But inequality (38) holds for all $t \geq 0$. Hence the inflexion point criterion is satisfied.

We shall now compute the slope of $h(x)$ at the inflexion points and inquire whether it increases with time. Substituting the expressions for $E(x)$ and $E'(x)$ given by equations (23) and (24) into (26) we obtain

$$\frac{\partial \bar{h}}{\partial x} = (I - h_0)ate^{a\frac{1}{1+x^2}t}\left[\frac{2x}{(1+x^2)^2}\right]. \tag{39}$$

Therefore, the slope of $h(x)$ at the inflexion point $x = x^*$ is given by

$$\left.\frac{\partial \bar{h}}{\partial x}\right|_{x=x^*} = (I-\bar{h}_0)\,\alpha t e^{\alpha \frac{1}{1+(x^*)^2}}\left[\frac{2(x^*)}{(1+(x^*)^2)^2}\right]. \qquad (40)$$

The expression for x^* is given by equation (34). One could, therefore, proceed to evaluate the slope at the inflexion point by substituting the right-hand member of (34) into the right-hand member of (40). This, however, would lead to extremely unwieldy expressions. An alternative approach is to solve for αt from equation (33) in terms of x^* and to substitute the resulting expression into equation (40). Our previous result on x^* indicates that as t increases, x^* decreases. Therefore, to ask how $\partial h/\partial x^*$ behaves as t increases is equivalent to asking how it behaves as x^* decreases, if all we are interested in is direction of change.

Solving equation (33) for αt we obtain

$$\alpha t = \frac{1-2x^2-3x^4}{2x^2}. \qquad (41)$$

Expression (40) becomes

$$\left.\frac{\partial \bar{h}}{\partial x}\right|_{x=x^*} = (I-\bar{h}_0)\,\frac{1-2(x^*)^2-3(x^*)^4}{x^*(1+(x^*)^2)^2} \times \operatorname{Exp}\left\{\frac{1-2(x^*)^2-3(x^*)^4}{2(x^*)^2[1+(x^*)^2]}\right\}. \qquad (42)$$

Note that since αt and x are always positive in the context of our discussion, equation (41) implies that $1-2(x^*)^2-3(x^*)^4 > 0$ for all t. This, in turn, implies that there exists an upper bound on x^*. It may appear that this upper bound does not depend upon the units of x. However, the units of x were fixed by the expression for $E(x)$.

It can be seen by inspection of equation (42) that both the rational and the exponential factors in the right-hand member increase as x^* increases. But we have previously shown that x^* decreases as t increases. Therefore, the slope at the inflexion point increases for increasing t.

The foregoing criteria for the progressive refinement of response, namely, the increasing curvature at the minimum threshold, the decreasing distance between the inflexion points of the threshold curve are, of course, not exhaustive. One might, for example, compare the minimum threshold with the average threshold as time pro-

gresses. Efficient learning would be characterized by a much more rapid change of the minimum threshold as compared with that of the average.

In the foregoing analysis we assumed that the threshold decreased linearly with the activity of the neuron. One can, of course, make other assumptions such as exponential decay. Thus

$$\bar{h}(x, t) = \bar{h}_0 \operatorname{Exp}\{-E(x)N(x, t)\}. \tag{43}$$

As in the preceding case we have

$$N(x, t) = \int_0^t [I(x) - \bar{h}(x, t)]\, dt.$$

This leads to the differential equation

$$d\bar{h}/dt = -EI\bar{h} + E\bar{h}^2 \tag{44}$$

and to the solution

$$\bar{h} = \frac{I\bar{h}_0}{\bar{h}_0 + (I - \bar{h}_0)e^{EIt}}. \tag{45}$$

As in the first case, $\bar{h}$ decreases in time and its rate of decrease is most rapid where E is largest.

Another modification can be introduced into the picture by assuming that the thresholds have a tendency to return to their original values. As a first approximation we may suppose that this tendency is proportional to the difference between the original value of the threshold and its mean value at a given moment. The differential equation for this case would be

$$d\bar{h}/dt = d\bar{h}_1/dt + \alpha(\bar{h}_0 - \bar{h}), \tag{46}$$

where $d\bar{h}_1/dt = -EI\bar{h} + E\bar{h}^2$ as in the preceding case. Equation (46) leads to the solution

$$\bar{h} = \alpha\bar{h}_0 + e^{-\alpha t}\int_0^t e^{\alpha t}F(t)\, dt - e^{-\alpha t}\alpha\bar{h}_0 + \bar{h}_0 e^{-\alpha t}, \tag{47}$$

where

$$F(t) = -\frac{EI^2\bar{h}_0(I - \bar{h}_0)e^{EIt}}{[\bar{h}_0 + (I - \bar{h}_0)e^{EIt}]^2}. \tag{48}$$

Generalizations of the Model.

The foregoing model depended upon the responses of a "success center" to the removal of noxious stimuli. This model can be easily generalized to one where both "pleasant" and "unpleasant" stimuli are taken into consideration.

Let now the linear array be innervated by two groups of receptors responding respectively to pleasant stimuli S_p and unpleasant stimuli S_u. As in the previous model, tracts from S_u inhibit the success center. The tracts from S_p, however, excite the success center. Now if some motor activity stimulated by some neurons in the linear array manipulates the environment in such a way as to increase the intensity of a pleasant stimulus S_p, this will result in a greater activity in the success center. This in turn lowers the thresholds of the motor neurons in the linear array which were involved in the "successful" activity. Thus by the same reasoning as in the foregoing case the learning of "successful" responses to pleasant stimuli is accounted for.

An obvious shortcoming of the models discussed here is the fact that in all of them not only the threshold of the "correct" neurons decrease, but, also, the thresholds of all the "wrong" neurons. This difficulty can be overcome by a mathematical artifact by allowing $E(x)$ to assume negative values. For such loci, x, where $E(x)$ is negative, we would then have the threshold increase with time as the inspection of equation (22) shows. This artifact is not applicable in our case. Our model is based on the supposition that the activity of the success center always tends to lower the thresholds in the linear array. Negative $E(x)$, therefore, would have no physical meaning, as a glance at equation (17) shows. The objection can nevertheless be met by the following considerations.

The rate of decrease of the thresholds of those neurons involved in the most "successful" activity will be the most rapid as the foregoing mathematical analysis has shown. Hence the average threshold $\bar{h}$ of the "best" neurons will reach S, the lowest level permitted by our equations, before any of the others. This imposed lower bound on $\bar{h}$, however, was dictated solely by convenience. It was made to avoid cumbersome discontinuities which would otherwise appear in equations (16) and (18). We can, however, extrapolate the results of our mathematical analysis by a qualitative argument.

Once the average threshold of a neuron in the linear array reaches the stimulus level, it may be expected to fire almost every time the stimulus is presented, provided the latter is of sufficient

duration. Its threshold will thereby be lowered even further. Its responses to the stimulus will become even more certain until a situation will prevail where only rare fluctuations of the threshold will be of sufficient magnitude and proper direction to *prevent* its firing.

In the meantime another effect will have entered the picture. While the average thresholds were still sufficiently high, the response-time included the expected time of occurrence of a sufficiently large fluctuation of threshold in the responding motor neuron. This means that in the beginning of the learning process the response times are, on the average, comparatively long. When, however, the thresholds of the "best" neurons, as described above, reach very low levels *their* response-times become very brief. This practically eliminates the responses of the other neurons since the response of the best neurons almost immediately eliminates the noxious stimuli. Thus the threshold distribution in the linear array becomes practically static as soon as the organism has learned to respond specifically and invariably to a noxious stimulus. For the case of pleasant stimuli discussed in the generalized model, this argument is not valid. This is so because the pleasant stimuli are not removed but, on the contrary, increased by the responses and thus may bring additional neurons into play.

A further generalization involves a "failure" center as well as the success center. Inhibitory tracts from the failure center innervate the linear array in much the same way as the success center. The effect of these inhibiting fibers, however, is to *raise* the thresholds of recently active neurons. The failure center is assumed to be excited by tracts leading from the receptors of noxious stimuli and inhibited by tracts leading from the receptors of pleasant stimuli. The function $E(x)$ will now be assumed to take on both positive and negative values. The results of the activity of a neuron at x affect the failure center in a way opposite from their effect on the success center, i.e. the strongest effect is by those neurons where $E(x)$ is minimal.

Let us now consider what happens at the failure center as a result of the activity of a neuron at x. If the response is "bad" i.e. $E(x)$ is small (or negatively large), the failure center will be stimulated most. Its activity will then result in the greatest raising of the threshold of the "offending" neuron. Hence future responses of that neuron will tend to be eliminated. The effects of "good" neurons on F are negligible and hence their threshold will not tend to be significantly raised but will be lowered by the success center as in the previous case.

The inevitable general lowering of thresholds in the first model may be also avoided by supposing the existence of an additional mechanism which tends to raise all thresholds indiscriminately.

Further Implications.

Suppose that the organism has been subjected to a given learning situation. As we have seen, the effect of such an experience would be to change its h curve.

If now the organism is subjected to a new learning situation involving the same learning mechanism, its new h curve will now play the part of the original $h_0(x)$ curve. Furthermore, the $E(x)$ associated with the new situation will, in general, be different. The question now arises: What are the invariants of such learning processes? One of the goals of the mathematical biology of nervous systems is to interpret such rather hazy terms as flexibility, memory, adaptability, learning transference, intelligence, abstraction ability, etc. in terms of parameters such as those discussed here. The notion of *special ability,* for example, to which we referred in connection with the "ganglion-brain," was interpreted in terms of biases in the connectivity of the net. We see that in the linear array model this notion assumes a somewhat more definite meaning. It may specifically refer to the distribution of connections from the success center to the linear array.

Alternative Interpretations of the Model.

It should be emphasized once again that our linear array is a mere mathematical artifact. Its only purpose is to "order" the neurons governing various responses in accordance with the effectiveness of the response. In particular, the loci of the array, which are in no way to be interpreted as geographical loci, may be considered as groups of neurons governing various *patterns* of response. These patterns may be simultaneous configurations of active motor organs or, even more generally, temporal sequences of such configurations. The flexion of a limb, the act of speaking, or even sequences of "thought" may be regarded as examples. The learning process then, according to this model, appears as a selective narrowing down to the optimal pattern of response.

The author wishes to express his appreciation to the members of the Committee on Mathematical Biology and in particular to Professors N. Rashevsky and A. Rapoport for their generous and valuable assistance in the preparation of this thesis. The author also

wishes to express his appreciation to Mrs. Carlton Smith and Miss Anna Wright for the pains they have taken in the preparation of this manuscript.

LITERATURE

Ashby, W. R. 1947. "Dynamics of the Cerebral Cortex Automatic Development of Equilibrium in Self-Organizing Systems." *Psychometrika*, **12**, 135–140.

——— 1948. "Design for a Brain." *Electronic Eng.*, December.

Householder, A. S. and H. D. Landahl. 1945. *Mathematical Biophysics of the Central Nervous System.* Bloomington (Indiana): Principia Press.

Lorente de Nò, R. 1934. "Studies on the Structure of the Cerebral Cortex. I. The Area Entorhinalis." *Jour. Psy. Neurol.*, **45**, 381–438.

——— 1934. "Studies on the Structure of the Cerebral Cortex. II. Continuation of the Study of the Ammonic System." *Jour. Psy. Neurol.*, **46**, 113–177.

Markoff, A. 1913. *Ischislenie Veroiatnostiei.* St. Petersburg.

McCulloch, W. S. and W. Pitts. 1943. "A Logical Calculus of the Ideas Immanent in Nervous Activity." *Bull. Math. Biophysics*, **5**, 115–133.

Rapoport, A. 1948. "Cycle Distributions in Random Nets." *Bull. Math. Biophysics*, **10**, 145–157.

——— and A. Shimbel. 1948. "Steady States in Random Nets: I." *Bull. Math. Biophysics*, **10**, 211–220.

——— 1948. "Steady States in Random Nets: II." *Bull. Math. Biophysics*, **10**, 221–226.

——— 1950a. "Contribution to the Probabilistic Theory of Neural Nets: I. Randomization of Refractory Periods and of Stimulus Intervals." *Bull. Math. Biophysics*, **12**, 109–120.

——— 1950b. "Contribution to the Probabilistic Theory of Neural Nets: II. Facilitation and Threshold Phenomena." *Bull. Math. Biophysics*, **12**, 187–197.

——— 1950. "Contributions to the Probabilistic Theory of Neural Nets: III. *Bull. Math. Biophysics*, **12**. Forthcoming.

——— 1950. "Contribution to the Probabilistic Theory of Neural Nets: IV. Simple Summation and Bilateral Inhibition." *Bull. Math. Biophysics*, **12**, Forthcoming.

Rashevsky, N. 1938. *Mathematical Biophysics.* Chicago: University of Chicago Press.

——— 1948. *Mathematical Biophysics.* Revised Edition. Chicago: University of Chicago Press.

Roberts, J. B. 1948. "Mathematical Biophysics of Some Neural Nets." *Bull. Math. Biophysics*, **10**, 1–10.

——— 1949. "The Group Structure of Some Neural Nets." *Bull. Math. Biophysics*, **11**, 51–58.

Shannon, Claude E. and Warren Weaver. 1949. *The Mathematical Theory of Communication.* Urbana (Illinois): The University of Illinois Press.

Shimbel, Alfonso and A. Rapoport. 1948. "A Statistical Approach to the Theory of the Central Nervous System." *Bull. Math. Biophysics*, **10**, 41–55.

Shimbel, Alfonso. 1948. "An Analysis of Theoretical Systems of Differentiating Nervous Tissue." *Bull. Math. Biophysics*, **10**, 131–143.

——— 1949. "Input-Output Problems in Simple Nerve-Ganglion Systems." *Bull. Math. Biophysics*, **11**, 165–171.

Tryon, R. C. 1931. "Studies in Individual Differences in Maze Ability: II. The Determination of Individual Differences by Age, Weight, Sex and Pigmentation." *Jour. Comp. Psych.*, **12**, 1–22.

——— 1931. "Studies in Individual Differences in Maze Ability: III. The Community of Function between Two Maze Abilities." *Jour. Comp. Psych.*, **12**, 95–116.

Uspenski, J. V. 1937. *Introduction to Mathematical Probability*. New York: McGraw Hill.

Wiener, Norbert. 1948. *Cybernetics*. New York: John Wiley and Sons, Inc. Paris: Hermann et Cie.